

COURTESY
OF CODE

CA20N
DE 816
-78 B07

INFORMATION BULLETIN NO. 7



SELECTION OF APPLICANTS
TO A FACULTY OF EDUCATION:
A CASE STUDY

W.S. PERUNIAK
M.W. WAHLSTROM
E.L. WEINSTEIN

MAY, 1978

COMMISSION ON DECLINING SCHOOL ENROLMENTS IN ONTARIO (CODE)

R.W.B. JACKSON
COMMISSIONER

HOWARD B. HENDERSON
EXECUTIVE SECRETARY

252 Bloor Street West
Toronto, Ontario
M5S 1V6

SELECTION OF APPLICANTS
TO A FACULTY OF EDUCATION:
A CASE STUDY

W.S. Peruniak
M.W. Wahlstrom
E.L. Weinstein

This paper was commissioned by and prepared for the Commission on Declining School Enrolments in Ontario and is not to be cited or quoted without the permission of the Commission and the authors.

This study reflects the views of the authors and not necessarily those of the Commission or the Ministry of Education.




TABLE OF CONTENTS

I	INTRODUCTION	1
II	HISTORICAL CONTEXT	3
III	REVIEW OF SELECTION ISSUES	8
IV	GENESIS OF THE RATING INSTRUMENT	28
V	INTERVIEW PROCEDURES	38
VI	ANALYSIS OF THE DATA	43
VII	RESULTS AND CONCLUSIONS	63

REFERENCES

APPENDICES



Digitized by the Internet Archive
in 2024 with funding from
University of Toronto

<https://archive.org/details/31761118906742>

CHAPTER I

INTRODUCTION

The main purpose of this study was to examine the Faculty of Education, Queen's University, data file of 840 applicants selected to be interviewed for admission to the B.Ed. program which resulted in a 1971-72 class of 640. A follow-up of the 640 students was conducted to determine how many had not received certification, who had accepted teaching positions and were currently employed by an Ontario school board in 1977-78, and which candidates were not employed as teachers in Ontario. Because an extended and a structured methodological procedure was used in selecting the 1971-72 class, based upon previous admissions procedures of Queen's and information from the research literature, it seemed appropriate to use this particular class for follow-up purposes to determine the relationship between data obtained in the admissions process and information about their current role in the Ontario educational system.

Data tapes from the Ministry of Education provided basic information on who was teaching during the school year 1976-77, as up-dated in files effective June 1977. When this study began it was not possible to obtain current data for the 1977-78 school term from Ministry of Education computer files about place of employment and a given teacher. The June 1977 information yielded a list of 557 certified teachers who had graduated from the 1971-72 class, of which 346 were currently employed, and was used to determine place of employment. Information about the remaining 83 students was provided by officials at Queen's University. The majority of the 83 students were not residents of either Ontario or other provinces and had not been included in this particular Ministry of Education computer file. Principals from 255 schools were contacted via telephone to provide information about the 346 teachers. This was done according to a structured interview schedule that provided data that was related to information obtained during the 1971 admissions interview and the 1971-72 teaching experience of the Queen's B.Ed. program.

Information in this report deals largely with the Faculty of Education, Queen's University, experience in structuring an admissions procedure based upon their experience and information from the research literature. Some six years later, by means of this study, we examined the validity of the admissions measures and are assembling research data on the teachers who have remained in the teaching force over the past six years, as described by their respective principals. We acknowledge that this is a global assessment and may have personality characteristics involved in the descriptions, but this is a first step and, indeed, a needed one as supervisory officials usually are charged with personnel selection and assessment. Only now, both in the research literature and in our schools, is a focussed concern developing on how to assess teachers and teaching effectiveness. That is not to say there is a dearth of material on the topic; indeed, there is an abundance of material but it has little apparent utility for school system application.

In this report we outline, in case study format, the events at Queen's related to admissions issues and solutions. As well, Chapter III outlines the research literature on admission and selection issues and provides a summary of conclusions that are relevant in 1978. Most of the literature cited is of relatively high quality and describes issues clearly. A dominant problem remains, namely, the so-called criterion problem. Before we can gain significantly in the quality of admissions decisions, we must pay particular attention to defining the attributes of a "good teacher" in terms of those deemed successful or effective as teachers, if this is to be a dominant criterion in the admissions phase of the B.Ed. program. Although somewhat dated, results from Fishman's (1958) work entitled "Unsolved Criterion Problems in the Selection of College Students" provides a basis for further study in the area of admissions decisions.

We anticipate that the results and conclusions presented in the report will be but one of the first steps in Ontario for structuring further studies to determine the nature of effective teaching and to describe "good" characteristics desired for selected candidates by a faculty of education.

CHAPTER II

HISTORICAL CONTEXT

During the period of educational expansion in the 1960's, the major issue in the teacher supply question concerned recruitment. Increasingly, during the period of educational contraction in the 1970's, the problem of *teacher selection* is supplanting *teacher recruitment* as the central challenge confronting the teaching profession, government authorities, and universities. Almost every teacher training institution in the province has recently experienced the difficulty of selecting from a large pool of applicants the number allowable under fixed enrolment targets. As the school system continues to shrink in the years immediately ahead, faculties of education may be expected to reduce the number of places in their training programs still further. When this development is combined with a relatively constant number of university graduates looking for professional training, it is safe to forecast that faculties of education will face unprecedented difficulties in selecting, on supportable grounds, the small cadre of teacher trainees required by the school system.

The difficulty of meeting that challenge is compounded by the dearth of accumulated experience. Expertise acquired during the years of recruitment does not directly transfer to the different needs associated with rigorous screening. Faculties of education cannot simply revert to the historic university mode of selection and base their admission criteria on academic standing. Their sense of responsibility to the public and to the profession would preclude them from resorting to such a simplistic solution. Unlike many other professions, teaching involves more than a delivery of services; it involves an intense reciprocal process with learners that calls for teachers with relational and interactive skills as well as intellectual competence.

The primary purpose of this report is to describe a case study of a systematic admissions procedure used at Queen's University and,

thereby, to contribute to the pool of shared experience. Although the number of studies on admissions procedures number in the thousands, the incremental gain in knowledge from many studies is minimal. This study systematically builds upon previous experience and is now, in its sixth year, concerned with an evaluation of the procedures by a follow-up of selected candidates.

When the Faculty of Education, Queen's University (known then as McArthur College of Education) first opened for classes in September, 1968, the major question regarding admissions was whether enough teacher candidates could be attracted to give the new Faculty and the new program a measure of credibility. The vast majority of university graduates who were interested in becoming secondary school teachers were still opting to enter the profession via the emergency two-summer-sessions pattern. Ironically, the heavy enrolment in the Queen's summer program of teacher preparation (which had functioned for several years as a satellite operation of the Ontario College of Education) contributed to the low enrolment of the fledgling winter program. Even with the inducement of a Bachelor of Education degree (the first such degree to be offered in Ontario for pre-service education), Queen's attracted only 189 candidates for the B.Ed. degree program in its first year. Under these circumstances, the initial concerns regarding enrolment focussed more on promotion and recruitment than on selection.

Almost overnight the situation changed. Word spread that there were novel features to the Queen's program. Consequently, some 1,300 university graduates applied for admission to the second year of the Faculty's operation. The enrolment target for that year, 1969-70, however, had been fixed at 200, with little possibility for any appreciable expansion until additional student residences promised by the province were built. Thus it happened that during the very first year of its existence the Faculty was confronted by the inescapable necessity for selective admissions.

It took several months to grasp the full dimensions of the problem. Then, early in 1969, the Dean appointed an advisory committee on

admissions to formulate a set of guidelines for selection. (Under the interim Faculty board structure there was yet no standing committee on admissions.) This advisory committee established academic grade thresholds in the various curriculum options and recommended that those applicants who met the requirements (generally a high B average) be interviewed by panels consisting of two or three instructors. By means of that academic eligibility filter, the Faculty succeeded in reducing the pool of 1,300 applicants to about 500 who were invited for interview.

An assessment form was devised to focus interviewers' attention on particular characteristics; namely, enthusiasm for subject, clarity of conversation, vitality, commitment, openness, and maturity. A final summation section forced interviewers to assign a global rating on a 10-point scale (later reduced to six points) ranging from not suitable to outstanding. Applicants were then rank-ordered on the basis of interview scores within their respective curriculum areas. An actual enrolment of 220 was generated for the academic year, 1969-70.

Staff reactions to this first experience with selection procedures were predictably mixed. Nevertheless, the difficulty involved in finding a consensus for alternative procedures led to an extension of the existing arrangements for another year. During the spring and summer of 1970, McArthur staff (with student participation this time) again conducted personal interviews at a number of Ontario centres. Of a total 1,081 applicants, over 500 met the academic criteria and were eligible for interviewing. The number actually interviewed amounted to 494; of those 494 interviewees, only 22 were not offered acceptances (giving a rejection rate of under 4½%). After some offers of acceptance were declined, the net enrolment for the third academic year, 1970-71, was 337.

It was during that second year's experience with the selection procedures that faculty concerns began to crystalize. Some staff members criticized the practice of according such a high weighting to academic grades. Others were disturbed by various aspects of the interviewing arrangements: the lack of a research base for the rating

instrument, the subjectivity involved in the evaluation, and the total absence of training for the interviewers. By the spring of 1970 there was a prevalent feeling that the faculty should make the admissions policy "one of our serious research projects in the College", and, subsequently, the new Admissions and Standards Committee was charged with recommending a new admissions policy before January 31, 1971. Concurrent with the emergence of faculty doubts about admissions procedures, advice from the influential advisory committee (not to be confused with the early admissions committee) strongly supported retention of personal interviews.

It seems important to understand why the opinions of concerned faculty, knowledgeable members of the profession, and informed lay representatives were divided. The most plausible explanation relates to the general state of the art in defining and predicting teacher effectiveness. Without a general consensus on what constitutes exemplary teaching behaviour, it must be difficult even for reasonable people to reach agreement on which indicators have the greatest predictive utility. That central problem is further compounded by the pressures of time and numbers. Faculties of education lack the resources to conduct an in-depth study of each applicant. They require a technology or a set of procedures sufficiently parsimonious to assess thousands of applicants within a few months and to determine which ones are the best prospects for teaching. Given the diverse perceptions of the ideal teacher practitioner and given the indeterminacy of most empirical studies in that area, it is not surprising that the Queen's Faculty of Education exhibited strong differences of opinion every time the issue of teacher selection was debated. Rather, it is to their credit that the faculty have chosen (1) to state their differences openly, (2) to recognize that the mixture of concern and uncertainty frequently signals a significant question, and (3) to mobilize energy and goodwill for a systematic probe of the question.

At the same period a seemingly unrelated event occurred which was to provide the foundation for the subsequent investigation. On October 26, 1970, the Department of Educational Administration,

The Ontario Institute for Studies in Education, sponsored a one-day "Seminar on Teacher Selection Simulation Materials", featuring Professor Dale Bolton of the University of Washington. Although Bolton's approach was intended for use with certified teachers, the Queen's Coordinator of the B.Ed. program, W.S. Peruniak, the senior author, envisaged that the methodology could be adapted to the selection purposes of the faculty of education. Furthermore, it was reassuring to note that Bolton's work was grounded on the extensive study of teacher characteristics conducted by Ryans (1960). After examining Bolton's materials and reading the background research, the Coordinator concluded that Ryans' and Bolton's work represented an unusually powerful approach to personnel selection, being based on a logical progression from rigorous empirical research to theoretical formulation and then to applied practice.

CHAPTER III

REVIEW OF SELECTION ISSUES

An examination of research literature on selection soon leads one to consider classification and placement, which are types of decisions about individuals made by institutions.

Cronbach (1971) has argued that selection is a type of classification problem where one of the classes of assignment is rejection. According to Cronbach and Gleser (1965), classification is the broad term to include a variety of people-assignment problems. It includes the assignment of people to different programs when there are several different predictor variables and several different programs. The assignment of people to different programs along a single dimension is a more restricted procedure called placement.

The concern in this report is with the selection problem where decisions are either to reject or accept. Typically, in faculties of education, students have considerable freedom to choose their own set of courses within a given program rather than yielding to a placement procedure. If additional admissions requirements are imposed beyond university level regulations for individual programs, the prerequisites tend to be minimal and associated with specific knowledge and/or skills.

Arguments about the need and type of selection procedures are numerous and divergent. Each faculty of education in Ontario has a selection procedure based largely upon academic achievement in the institution previously attended. Institutional resources and the number of applicants largely dictate the extent of selection and admissions procedures. In most cases in Ontario, selection procedures are designed to select a quota of students from a group of applicants who generally meet the minimum requirements. The dominant criteria for selection of candidates appears to be prediction of success in the faculty of education academic program rather than the criterion of

remedial courses to supplement existing programs. When making admission decisions, an institution is faced with restricting itself to certain kinds of applicants where the number of applicants far exceeds the desired quota. If the number of applicants is relatively small in relation to the number desired for a particular program, the decision may be to recruit applicants. In recent years, applicants have far exceeded the number that can be admitted to programs. Given this situation, decisions are then required on what measures are desired for use in making admissions decisions, how these measures will be combined and weighted if more than one is used, and where the decision point(s) will be on whatever measure or measures are used. The single score, be it from one variable or a combination of variables, is then used to decide whether an applicant will be admitted or rejected.

Many policy decisions regarding the purpose of programs are assessed to determine who should be selected. Most academic institutions appear to operationalize their policy to select those candidates they predict will survive their programs of instruction and ultimately graduate. In some cases, institutions behave as though they believe that even brief exposure to university programs is beneficial to individual students, even if there is a high probability of failure. Where institutional resources are abundant, some appear to believe that all persons should be given an opportunity to attend the desired program even if success is remote.

Concerns expressed in the above statements are largely with the predicted probability of success. During the seventies, however, admissions policy has often reflected a matter of economics rather than education, as budgets reflected income that was directly proportional to the size of the student body. When economics were not a primary concern, the issue became the perceived "quality" of the graduate which would be reflected upon the institution as the students entered the world of work. In this situation, the deliberate intention is to change students into models of enlightenment, where the benefit to the institution is in the growth of the selected candidate. Whatever the

determination, in order for the institution to select intelligently, it must take a stand concerning the kind of benefit it wants to maximize on the average in its institutional selection decisions. Unfortunately, however, many institutions have not made deliberate and explicit decisions about this basic issue.

If the institution feels that its programs are fixed and determined by the graduate from a one-, four-, or five-year program, it will not offer remedial courses. Thus, staff would feel that it is unreasonable to attempt to accommodate itself to applicants who cannot be expected to accomplish the objective within the specified period of time. Such an institution would be unsympathetic to the idea of introducing remedial courses for students who were inadequately prepared, or introducing more courses in a sequence, resulting in a longer curriculum for those who could reach the objective but only through more time and effort than the optimum. The situation in this special case could relate, especially in Ontario, to those persons who appear to be academically competent but who have an inadequate grasp of the English language. Thus, one way to adapt the program to potential students would be to offer special courses in English as a second language.

The sources of applicants may pose a special problem. On the one hand, one could leave the matter to chance and consider whoever applies. However, faculties of education in Ontario are supported by public funds and may feel an obligation or pressure to serve, primarily, those students who come from their geographical region. In the case of Ontario, there is no apparent mandate or obligation to students of one religion or sex, although it is of some considerable interest to have equal opportunity for females. The Ontario Teacher Education Colleges also feel a pressure to ensure that a desired proportion of selected candidates are affiliated with the Roman Catholic Church, as they will staff the schools designated as separate schools or Roman Catholic schools. Beyond these considerations, there is also the issue of whether the institution will actively recruit candidates. From the point of view of decision theory, even when available selection tests are not very efficient, utility can be increased through improved recruiting

for a constant-sized student body. This permits use of a more stringent criterion of quality for the students who will be admitted (Cronbach and Gleser, 1965), and this lower selection ratio has the same effect on increasing average utility that a higher selection-procedure validity would produce.

The Faculty of Education, Queen's University may also use its public image as a selection device. In a setting where the reputation is high, admissions policy and selection machinery provide the direct selection, but the public relations office provides indirect selection by controlling the image of the institution so that students select themselves in ways desired by the faculty of education. In one sense, direct selection can be dispensed with if the indirect procedure results in recruiting the kind of students the faculty wants. There are some problems, however, with indirect selection. Firstly, the faculty cannot tell very much about its selection ratio. Secondly, images are inherently more resistant to change than are institutions. They often lag far behind the reality of the campus. An admissions procedure that does not depend heavily on self-selection is much more controllable.

Although vast experience rests with a registrar's office and many research studies have examined the measures used by Faculties of Education in making decisions, we seem to have reached a plateau regarding significant improvements in the use of various measures. However, a rational selection program rests upon the deliberate and informed choice of the material to use. A major consideration in choosing material for selection purposes is the amount of material to be gathered and processed directly, which affects the cost of making each selection decision. It is, indeed, an interesting exercise to determine when the amount and type of information received have no additional impact on selection and, hence, should not be included. The extreme to be avoided is when so much testing and information gathering is being conducted that the measurement program costs more than it is worth. Although one can analyse costs for various

components, it is a policy decision as to how much one should extend costs in relation to return from the investment.

After selection of measures is completed, the next stage is to determine how the measures will be combined into a form from which consistent decisions can be made on applicants. The regression model and the cutting-score model are in common use, with individual characteristics included by each institution. Each model has different implications and produces different results. However, the usual situation is that such contrasting models will select the dominant group in common and differences between selection and rejection of applicants by each model varies increasingly as the "cut-off" score from either model approaches the decision point of accept *vs* reject. With regard to faculties of education and the relatively small number of applicants, many informal procedures are usually part of selection decisions. It seems that selection decisions are more frequently made by using a combination of a clinical judgment and statistical data.

Following from the measures and an appropriate model to utilize in selection decisions is the concern with the location of the specific decision point or points on the admissions measure or measures--the point above which a candidate is selected and below which he is rejected. Many factors are considered in deciding upon an optimal point. A faculty of education may select this point on the basis of the desired number of students to graduate from this class. Other considerations could be on the basis of a predetermined level of minimal acceptable talent, or on the basis of a minimum probability of success that the faculty of education can expect from the selected group. Alternatively, in the case of the Ontario Teacher Education Colleges, the Minister of Education established the quota for the 1978-79 academic year. Similar action may result from policy bodies responsible for all universities within the province. In terms of educational selection being a form of placement, the deciding point or score can be the point at which rejection is more beneficial to society than acceptance would be.

In some selection models the decision point is very precise, whereas in others there are essentially three groups: clearly accept,

decision to be made, and clearly reject. Such a "decision to be made" or decision area can employ additional variables, or random factors may even be permitted to operate. It is with this group that, in fact, decisions are made. Because the decision area group is at the cutting line, it provides an opportunity to involve a special additional set of measurements in the form of a try-out program, such as being accepted on probation or for a special trial period. An interesting way to view acceptance of applicants within the decision-theory frame of reference is that all acceptances in the admissions process are really temporary in that admission does not guarantee graduation. The candidate may fail during term exams and subsequently leave, may be counselled out by staff, or may decide to leave for personal reasons.

An overview of the selection process was outlined above with brief mention of procedures in Ontario. The issues raised are common over institutions and only differ in their particular application. One source of information as to what has been useful is the research literature.

The measurement process entails asking about criteria and standards, where standard in this case is reflected in the acceptance-rejection decision. For faculty of education programs, one criterion is that of a "good teacher". However, given the nature of teaching, little agreement can be achieved on determining fine distinctions between the quality of teaching of two graduates. Thus, we are left with the situation where it is seldom possible to measure the criterion of interest directly. Faculties of education wish to balance their books in the dual responsibility to society and the student and, thus, tend to select grades in courses as an indication of success in a program. Whether grades obtained at university and ability to teach are directly related or not remains an open question having supporters on both sides of the issue.

There are literally thousands of studies of the various kinds of measures for predicting academic grades. Regardless of whether academic grades measure personality attributes, effort, and other personal characteristics, grades have been chosen as the criterion of

success in school and university. Regardless of what one reads or hears about academic grades, especially regarding the goals of a faculty of education, institutions behave as though they valued most the students who obtain high grades. Indeed, grades are readily available, quantifiable, and are of importance in such decisions as being able to graduate.

Although average grade is frequently used as a criterion in selection, it has been criticized on a number of grounds. Klein and Hart (1968) argue that grades are a poor representation of educational utility because they are contaminated with irrelevant factors such as diligence, handwriting, and general verbal ability, as well as a personal attractiveness and skill in interacting with the instructor. A second argument surrounds the variability of grading and of different grading systems. Interpretation of grades then becomes difficult over different institutions, their departments, and the general time in which the program was completed. In Ontario, for example, there is considerable concern over the fact that the secondary school grades have steadily become higher since 1967 when provincial examinations were abolished. Also, there has been a reduction in the number and type of required courses needed for graduation purposes in secondary schools and faculties of education. Hence, the lack of compatibility of programs taken by students has reduced the comparability of average grade from person to person and group to group. During recent years, it is frequently argued that the spread of grades has generally diminished and is so reduced that individual differences in average grade have become unreliable. Secondary school and undergraduate course grades have been criticized as being unreliable, but the general opinion of researchers who conducted the Ontario "Interface Study" is that grades remain a reasonable and stable indicator for the prediction of academic achievement. As a general conclusion, however, the empirical character and philosophical bases of grading practices at the secondary school and university level are implicit and unexamined, so that no one really knows how grades are being arrived at or how they relate to the goals of the institution. Although there is much rhetoric about achieving

consistent policy and grading practices, there appears to be little systematic attempt to relate grades to the goals of faculties of education, and there do not appear to be consistent grading practices related to the concept of utility.

Because of perceived difficulties with grading practices and expressed dissatisfaction, other individual or group data have occasionally been used. One source of additional data has been ratings by faculty members. Research with procedures in industry, civil service, and education have indicated strengths and especially weaknesses (Campbell and Fiske, 1959; Davis, 1964, 1966; Guildford, 1954; Brown, Weinstein, and Wahlstrom, 1978; and Weinstein, Brown, and Wahlstrom, 1976). When used in academic situations, interview ratings tend to be difficult to obtain, of low reliability, and highly redundant with grades. Various Ontario institutions such as the Queen's Faculty of Education and the Ontario Teacher Education Colleges have, in recent years, attempted to overcome some of the difficulties with interviews and have embarked upon a course to improve data from this source. Although some professional groups use certification by an external agency as a further means to distinguish between expected performance levels within the profession, the current certification procedures in Ontario are generally based upon information about graduation from an institution rather than upon an additional examination.

There has been a trend in education to initiate competency-based testing for certification of teachers, but this work is still at the experimental stage and has had virtually no influence upon certification or graduation requirements in Ontario. Some information is collected on application forms about accomplishments in non-academic areas, such as prior work experience, but the type of information is often unique to an individual and is entered in a clinical manner rather than a statistical analysis to help determine an applicant's suitability for entering a teaching pre-service program. In the special case of faculties of education, applicants frequently are young and have not had an opportunity to gain non-academic work experience. Accordingly, it is suggested that they should not be penalized for this lack of opportunity by making this

type of activity a requirement for admission. A further consideration could be persistence until graduation. That is, in making admissions decisions, there has been concern about no-shows and in having selected candidates drop out once admitted. To date there is minimal evidence as to who will be a no-show and who will drop out, but these are important considerations and are currently being investigated at the Queen's Faculty of Education and the Ontario Teacher Education Colleges.

A review of the research literature provides evidence about decisions to operate a fixed-treatment program or an adaptive-treatment program. Given that faculties of education tend to operate upon the assumption of a fixed-treatment program, and if we further assume a linear relationship between pay-off and predictor variables, then Cronbach and Gleser (1965) have shown that it is desirable to test at least twice as many bona fide applicants as will be accepted, if the predictor variables are worth using at all. However, this is not the case in an adaptive-treatment program. In practice, and independent of the type of program officially offered at an institution, faculty members as individuals frequently alter their programs to suit the people they find in their classes and, in essence, introduce a form of adaptation. The consequence for a faculty of education is that treatment of students within programs will have important bearings on the use of predictor variables and on the optimum selection ratio to maximize gain in utility to be obtained through considering predictors at all.

A faculty of education's sources of applicants are factors to consider in developing admissions programs. One can maximize the selection ratio with known information if applicants come from a variety of sources, but this procedure may be inefficient if a faculty serves a local area and admits nearly all applicants. If the latter is the case, funds spent on admissions purposes simply may be wasted. If an institution does recruiting, it is important that the nature of the institution not be misrepresented or exaggerated as this may increase the number of dropouts later.

A general consideration in using prediction measures is that of seeking the greatest gain in utility by using the least expensive

measures that yield comparable results. That is, if useful predictions can be made from data that are collected for other purposes and, therefore, add nothing additional to costs, such data have the potential of providing appreciable gain. Thus, one should consider the readily available data as predictors. The second recommendation is that predictors cannot be considered in isolation from each other. Each additional measure has value only to the extent that it adds validity beyond that provided by the initial predictor, and each additional measure must be evaluated by its ability to improve upon the existing team or by the incremental validity (Sechrest, 1963).

A transcript is usually available describing an applicant's success in previous programs. This readily available document provides a data base for prediction that has had a long history of success. Many studies have demonstrated that secondary school grades are good predictors and that secondary school achievement is the most valid single predictor of performance in university as reflected in the criterion of university grades. Such a finding is to be expected since secondary school activities are in essence a work sample and are similar in form to the work expected at universities where the criterion to be predicted, namely academic grades, is reflected in previous experience. When using measures for prediction in selection decisions, one must remember that measures can have a positive correlation because of similarity in form as well as similarity in content or to dependence upon common underlying attributes. Accordingly, ratings might tend to predict ratings better than test scores; test scores should be the best predictors of test scores and, by the same argument, grades should be the best predictors of grades, as typically they are. Thus, in establishing an academic selection system, the first kind of a predictor to look for is some readily available work sample from previous performance. The best work samples usually will be the ones closest to the criterion in time and form. If the Queen's Faculty of Education admissions committee is concerned with academic success in their program, which follows an undergraduate degree, one of the best predictors for Queen's surely is the previous academic record. However, the matter is

generally not as simple as this because considerable concern is with prediction of success in teaching. Because the predictors are now significantly different from the criterion, and what constitutes good teaching is ill-defined, the ability to achieve a desired level of predictability is less than the institution desires. By way of noting that this is an important issue, Queen's Faculty of Education has systematically assembled admissions data from the 1971-72 class and has followed up on those who remain in teaching to determine the relationship between data available during the students' program year and employment data after several years of teaching experience. However, when grades at university are the criteria for success in a program, one of the best predictors will be previous university grades. If there is a desire to improve on the level of prediction available from the conveniently accessible work sample, one should evaluate carefully the increment to validity that is obtained in using additional data.

When using academic records of performance from a variety of institutions, it is expected that different standards will have a differential effect upon predictions of university success. Such was not the case in using unadjusted secondary school grades and test scores as predictors. That there is no demonstrated influence was reported by Lindquist (1963) and Linn (1966). The rationale is that test scores serve in a regression equation to counteract the influence of differential grading standards in secondary schools where, in effect, they equate the secondary school grades from various schools. Evidence such as this supports a recommendation for introducing a testing program for admissions purposes when applicants come from a variety of institutions having different procedures and standards for assigning grades. Just as secondary school records and aptitude test scores make an excellent prediction system (Hills, Gladney, and Klock, 1967), so should university transcripts and scores from a testing program. Whereas the practice of rank-in-class has decreased in Ontario schools, rank-in-class and secondary school average serve about equally well as predictors, especially when they are used in conjunction with other predictors such as admissions test scores (Hills, 1971). The

distinct advantage is that schools do not need to wait for students to complete a grade or to graduate before a reasonably complete representation of secondary school performance can be made. A marked difficulty arises, however, in the case of large secondary schools, as the procedure is only optimally effective if all students are ranked.

In addition to providing a transcript, an applicant generally completes an application form requesting personal data. Although the data may be used for specific purposes, such as to identify persons with specific talents, this source of information for academic selection often lies untapped. Opinions and results are mixed regarding the value of such data and some methodological problems are often overlooked in studies using biographical data. Hilton and Myers (1967) have noted that incremental validity of a new predictor is important if it is to be of practical significance. As well, any variable must meet the conditions imposed by cross-validation where the stability of a variable's contribution to the prediction is assessed. In the special case of biographical data, where selected items have been determined empirically from an initial sample, cross-validation is especially crucial.

Although cross-validation is important, it places second to operational validity being diminished. Knowledge about tests, admissions procedures, and prior decisions based upon such information may affect the way persons apply and may even affect who is willing and interested in completing an application.

One of the difficulties of using biographical data is that what works in one situation may not work in another setting. Thus, it is difficult to recommend any specific set of such data that might be relevant to a given situation. A further concern is that there is no theoretical framework for dealing with the structure and associated data base.

A methodological issue of considerable concern is when the effects of prior selection are not included in studies that report satisfactory incremental validity. When previous grades and test scores have been used in the selection of candidates on whom the validity of biographical

data was examined, unused variables will tend to appear superior in predictive efficiency due to the restricted range of the variable already used in selection. Detailed reports by Horst (1954, 1955) and Lunneborg and Lunneborg (1966) outline absolute and differential prediction settings utilizing biographical data as predictors. Only limited facets of using biographical data have been investigated although hundreds of studies have been reported. Studies tend to duplicate the work already done and there is a need for extensive methodological developments prior to any breakthrough in using such data.

Use of biographical data as non-intellective predictors appears enticing at first glance, if one attempts to relate current non-intellective factors to a future attainment in this area, because it is an additive component to the largely academic prediction model. However, lack of conclusive evidence for such variables, together with obvious methodological flaws in several studies using this type of data base, makes one cautious about a positive recommendation for routine use in selection decisions. At this point in time, it is also difficult to suggest any meaningful path to explore that should lead to success.

Whereas studies on biographical data for prediction purposes have largely been negative, evidence from work on previous academic performance and scores on admissions tests have often been found to be useful. Typically, admissions tests are closely related to success criteria and, hence, are deemed to be extremely relevant for admissions purposes. One of the most common procedures is to use a test battery containing verbal and mathematical aptitude subtests. It is common knowledge that these two sets of scores are optimal areas for deriving a measure of intellectual competence associated with performance in academic programs. Horst (1966) reports that combining previous academic performance and these two scores results in a large proportion of criterion variance being accounted for in prediction equations. Although reported studies frequently cite use of either the Scholastic Aptitude Test of the College Entrance Board or the American College Testing Program, it is suggested that similar incremental validity can be obtained when using any of a wide variety of aptitude tests commercially available. If one wishes to add

achievement test scores to those of previous academic performance and aptitude test scores, one will generally find a modest or inconsequential gain in multiple prediction models. A partial explanation is that there is not a clear distinction between aptitude and achievement test results of young adults. That is, it is difficult to measure a university graduate's intelligence uncontaminated by previous academic experience and training. Prediction equations using such data are generally adequate for use in immediately succeeding years, but one should check the beta weights after two or three years to ensure continued operational validity as the relative weights may shift due to different emphases in previous academic performance and the current program at a particular university.

Evidence from the research literature suggests that a sound educational selection system should include an evaluation of past performance, such as reflected in grades, and, secondly, have added to this evidence one or more scores from tests of academic aptitude and achievement. Addition of procedures and data beyond that described have been tried with little consistent evidence of incremental validity. As noted, biographical data has been used with varying degrees of success. A very common procedure that apparently has high face validity in faculty of education admissions is to have each applicant interviewed by a representative of the institution. Evidence on the value of interview findings as incremental predictors is very discouraging and has been for years (Cronbach and Gleser, 1965, Kelly, 1954; Weinstein, Brown, and Wahlstrom, 1976). It seems that because we have so much personal faith in interviews we are reluctant to accept evidence to suggest that such procedures are inadequate. Consequently, considerable effort has been devoted to an analysis of the interview process and its role in selection decision. One speculation is that the so-called unstructured interview is relatively lacking in efficiency when compared to the structured interview. In essence, the structured interview is similar to a test and takes on many of the positive attributes of testing which would suggest that results from a structured interview for a selection process would illustrate a significant improvement over

those from unstructured settings. A tentative recommendation is that one should provide in-service training for interviewers and that interviews be conducted using a structured rating system rather than a general discussion approach. General findings from studies of the interview process are that (a) decisions that are made in the selection interview are usually made within the first 30 seconds of the interview, even though an average interview time period is 15 minutes, (b) early impressions are very important, (c) the interview tends to be a search for negative evidence, where anything unfavourable that appears in the interview is likely to lead to rejection, (d) interviews often become sales pitches during the latter minutes when the interviewer has already decided to recommend acceptance of the applicant, but if rejection is the decision, the interviewer is less pleasant, and (e) interviewers tend to differ in their response styles, particularly in their category widths, i.e., their tolerance in accepting candidates. Because of the unstructured nature of many interviews, personal idiosyncrasies of the individual are allowed to influence interview ratings and, hence, may be a major explanation for the differences between the decisions of different interviewers. When considering the final decisions that will be made from interviews, it appears that unfavourable impressions are much more important than favourable impressions. As in the marking of essays, interviewers are sensitive to adaptation level where a poor applicant tends to make the applicant who follows look good and, conversely, a good applicant handicaps the person following. One measurement perspective is that increased structure and greater concern and control over the interview content would see a significant improvement in interview results. Whereas the focus tends to be upon reliability and content validity, the greatest need is to establish construct validity in interviews for incremental predictive purposes. Implicit in these statements is the question of the most relevant criterion. If one is selecting on the basis of academic performance as the criterion, then previous academic grade average should be one of the most useful indicators. However, if one wishes to use a different criterion, such as high quality teaching performance, then an interview may yield "good"

results. It is here argued that the interview can be useful as a selection tool, especially if greater attention is paid to psychometric considerations. However, at present the contribution of the interview is largely beneficial as an in-service program for faculty and to public relations. It appears that few negative categories arise with the interview apart from time and cost considerations.

Cronbach and Gleser (1965) have argued that application of wideband procedures may justify continued use of interviews. The principle is that information of modest validity used in many decisions may be more beneficial than information that is highly relevant for but one decision. A dominant feature of the interview is that it is a viable procedure for obtaining information on a variety of topics and great variety can be introduced without increasing the cost proportionately. By means of such an argument, one could justify an interview over a long admissions test relevant for predicting academic performance. The weak point in such a statement is that interview information must be systematically disseminated and used if it is to be beneficial. Perhaps more of the unanticipated gain from interviews occurs through the in-service phenomenon that is fully appreciated. From an economic and efficiency perspective, the interview has low utility for most admissions decisions. However, from the writers' perspective, it would be most interesting to analyze institutional actual dollar expenditures in situations using interviews and not using interviews. Also, one would wish to examine the productivity gain within a faculty if their time were devoted to other activities. Although such a study may be difficult to conduct and it may not be easy to quantify the different sets of outputs, such an analysis may reveal greater benefits from interview procedures than is normally claimed.

References and ratings of qualifications or personal characteristics are routinely required by many educational institutions and in many ways are similar to the interview. There is minimal evidence in the research literature on the effectiveness of such data. As such procedures are generally of a heterogeneous nature and do not specify the characteristics

being rated, there is no empirical basis for expecting ratings to provide appreciable incremental validity. Gough, Hall, and Harris (1963) and Smith (1966) report disappointing results in the effectiveness of such data for admissions decisions.

Just as there has been a glimmer of hope associated with increased use of biographical data and interviews, much effort has been expended upon incorporating personality variables into selection procedures. Chauncey and Frederikser (1951) suggested that the greatest advances may come through a thorough exploration of the measurement of personal qualities. Little has apparently been accomplished since that early statement was written. Smith (1966) has cautioned us against optimism and it is difficult to see where any one of the several lines of effort has moved us forward over the past two decades. It seems logical, especially in the domain of teaching, that temperament and personality should play an important role in academic performance and should indeed influence how a person performs as a teacher. However, there is little encouraging evidence to report. According to Lavin (1965), there is no sign of the "big breakthrough" here, despite a great amount of effort being devoted to the problem. An explanation offered is that many of the criteria used in selection may share elements of personality measurement and, hence, do not contribute a significant variance component in prediction equations when used as separate scores or variables.

Different studies incorporate varying terminology, such as the use of work samples for prediction purposes. For purposes of a data base to be used by faculties of education, the work sample is essentially represented in the secondary school and university academic record and is, perhaps, the best work sample possible. Persons using academic records should remember that three- and four-year university programs, summarized in a one-page transcript, are massive data bases and should not be considered lightly in any decision process. In fairness to admissions officers, however, we acknowledge that data other than previous academic performance is desired in order to have a broad data base.

By way of a summary of non-cognitive predictors, it seems that only minimal gains in prediction above and beyond what is attainable via the

use of intellectual predictors alone have been demonstrated. We continue to seek an expanded data base for prediction purposes, and continually argue that components other than the academic are relevant, but our success in this area has been less than our rhetoric would suggest.

The review of issues related to the selection problem has been largely based upon the work of Hills (1971) and is freely adapted for this report. Because the number of studies is literally in the thousands and many excellent reviews are available, it was deemed appropriate that a modification of existing material would be more valuable than another review at this time. Thus, we gratefully acknowledge the extensive use made of Hills' reports and suggest that the material presented here is more comprehensive given this particular approach.

The review of materials presented dealt with the type and nature of variables usually considered in admissions decisions. Once the variables have been chosen, however, the core of the process begins. Now the admissions office must decide which students will be accepted and which rejected. These selection decisions are usually made by one of three common methods: (1) allow an admissions officer to study the applicant's record, weight the evidence as clinical judgment dictates, and arrive at an admit or reject decision; (2) establish cutting scores on each variable as a basis for the admit/reject decisions; or (3) combine two or more variables into a single score, usually by a linear combination, and admit the applicants with the highest composite score.

The first alternative, clinical judgment, is probably the most widely used selection method. It uses human judgment to integrate applicant data, as well as to select the variables for consideration. It allows consideration of each applicant as an individual case. It avoids the time-consuming calculations common to many mathematical decision-making techniques. To its detriment, however, it has been repeatedly demonstrated that mathematical methods make more accurate decisions than admissions personnel. Clinical judgment is also criticized because it does not allow decisions to be made openly by explicit and defensible criteria. Thus, clinical decisions tend to obscure institutional selection goals

and promote the quality standards of the individual decision maker.

The second alternative, use of multiple cutting scores, has two forms. In its most basic form, the institution sets cutoff scores on each variable and then selects those with scores above all cutoffs. Quite often cutoff scores are set arbitrarily and yield an unexpected number of enrollees. There are several methods, however, for setting cutoff scores properly. These methods may take account of the cost of gathering information about the applicant, the validity and reliability of information obtained, and the desired enrolment of each program. Cutoff determination procedures can also be used in the more complex model. In that model, two cutoffs are set for each variable. Those below the first cutoff are rejected, those above the second cutoff are accepted, and decisions about the remaining group are made by clinical judgment or linear combination. This technique allows institutions to quickly identify and accept the most desirable students, and permits establishment of minimal standards for admission. It also helps cut admissions costs while preserving or enhancing the quality of decisions made. Multiple cutoffs are easy to compute, to administer, and to explain to laymen. The separate scores involved provide a valuable basis for guidance counselling. The strict cutoff approach assumes that compensation between needed skills is not feasible. With the two-cut-off model, this assumption is tempered by the assumptions of the method used to deal with the undecided group. It is preferable that the assumption of noncompensation be moderated, since experience suggests that in education, and especially in institutions with varied curricula, lack of one kind of talent can often be compensated for by possession of another relevant talent.

The third method, linear combination, also has several forms. Nonetheless, one may say that linear combination assumes skill compensation. That is, a high aptitude of one relevant kind can compensate for lower aptitude in another relevant skill. Linear combination methods, in general, have relied upon computers to provide highly sophisticated multiple regression equations as a basis for a composite score. These methods provide exact solutions in a statistical sense, but generally

one finds that the regression weights can be simplified dramatically with little practical loss of predictive accuracy via "hand" or non-machine solutions. Multiple regression, either with a sophisticated computer system or via a nominal method readily derived using routine office procedures, is recommended in an appropriate form. Technical advice, however, is also recommended when an admissions office is using such a system, as both statistical and measurement problems frequently are not recognized by admissions personnel and thus allow unintended assumptions to be incorporated into the admissions model. A trivial example is that different means and variances for each variable can serve as unintended weighting factors and have an undesirable effect upon decisions regarding acceptance and rejection of applicants, especially for those near the cutoff score. It should always be remembered that all of these selection methods are limited by the quality of the predictor variables which they use. Furthermore, the decision-making procedure should be chosen which can best serve institutional goals -- the tail shouldn't be allowed to wag the dog.

A frequent practice when various kinds of data such as test scores and biographical data are assembled is that a human judge makes a decision as to which candidates will be rejected. Such a person may have access to statistical summaries and the candidate's file and, although the issues are not resolved fully, it seems clear that the mechanical statistical combination of data is practically never improved upon by modification by human judgment (Sawyer, 1966). However, if judges or interviewers are to be used in the selection process, present findings suggest that the data generated be introduced into prediction equations or other statistical combination procedures that generate the final decision.

Persons interested in implementing an admissions procedure will need to examine the various kinds of benefit, criteria, and possible measures for making decisions. Accordingly, this will lead to the setting of cutoff scores that may be examined from the perspective of classical (predicting categories for measurements) and decision theory (gain in utility) procedures.

CHAPTER IV

GENESIS OF THE RATING INSTRUMENT

The theoretical foundation for the interview-rating instrument was outlined in *Characteristics of Teachers*, where David G. Ryans states that the Teacher Characteristics Study, which he headed, represented "one of the most extensive research programs that has been directed at the objective study of teachers" (p. 6). Over a period of six years (1948-54), approximately 100 separate research projects were carried out, involving more than 6,000 teachers and 1,700 schools in 450 school systems across the United States.

One of the principal concerns that prompted the research was "the need for procedures for appraising certain characteristics of prospective teachers before or during pre-service training and at the time of employment by school systems to help improve teacher selection and assignment" (p. 9).

A major objective of the study, growing out of that need, is described as follows: "The identification and analysis of some of the patterns of classroom behaviour, attitudes, viewpoints, and intellectual and emotional qualities which may characterize teachers" (p. 9).

The general procedure of the study consisted of the following phases: (1) extensive observation of the classroom performance of teachers, resulting in a collection of more than 500 "critical incidents" or significant teacher acts; (2) subsequent selection of relevant first-order teacher behaviour dimensions from that assembly of "critical incidents" for inclusion in new observational instruments; (3) the assessment of the classroom behaviour of large numbers of elementary and secondary school teachers, utilizing those instruments; and (4) statistical analysis of the teacher behaviour assessments leading to the identification of some major patterns of teacher behaviour, as observed in the classroom.

The assessment form, which came to be known as the Classroom Observation Record (C.O.R.), underwent numerous revisions. As mentioned

above, the behaviour dimensions chosen for inclusion in the C.O.R. were derived from the lists of significant teacher behaviours or "critical incidents." A basic methodological assumption in the assessment record's format was the hypothesis that personal-social traits may be conceptualized as bi-polar dimensions, the opposite poles of which can be described with reasonable precision. The final version of the record utilized a seven-interval scale for rating those dimensions.

It is instructive to note that the researchers attempted at an early stage "to reduce the assessment procedure to the tabulation of the frequency with which each of the specific behaviours was observed for a particular teacher (p. 84)." This mechanical scoring system, however, had to be abandoned in favour of a more impressionistic procedure. Ryans explains:

A number of scoring systems for such a check-list approach were derived and tried out, but the technique proved cumbersome and less reliable, as judged by interobserver correlations, than the earlier-used estimation procedure. The apparent objectivity of a check-list approach makes it particularly attractive, and the results obtained in this type of attempt to assess teacher behaviour was (sic) disappointing to the staff of the Study. Use of check-lists, therefore, was discarded in favour of the more intuitive procedure, standardized and controlled through use of the Glossary and by training of the observers, but nevertheless relying upon a less objective summing-up of specific behaviours in arriving at an assessment (pp. 84-85).

This excursion into methodological history was prompted by the unexpectedly high degree of inter-rater agreement revealed in the Queen's University procedures, this despite the lack of rigorous attention to ensuring standardized observer performance. It is at least conceivable that, in view of Ryans' indifferent success with categorical check-lists, there is some unrecognized basis for consistency in the intuitive approach to behaviour assessment when utilizing carefully defined bi-polar dimensions and when employing an alert, educated group of observers (even if technically untrained).

To return now to the general procedures of the Ryans study, once the data provided by observers' assessments had been gathered, two

independent factor analyses were undertaken -- one on the intercorrelations of assessments of 275 elementary school teachers, and the other on the intercorrelations of assessments of 249 secondary teachers.

With reference to the assessments of elementary teachers, Ryans writes (p. 97)

Product-moment correlation coefficients were computed among twenty-four of the dimensions. The resulting table of intercorrelations was factor-analyzed by the centroid method and both orthogonal and oblique rotations were attempted. Five centroid factors were extracted. Orthogonal rotation of these factors did not yield an acceptable solution. The oblique factors, however, provided a solution that more satisfactorily met the customary criteria of simple structure.

Following a similar procedure for secondary school teachers, Ryans extracted six factors from the 25 variables involved.

Three factors were common to both elementary and secondary teachers:

These primary teacher classroom behaviour patterns were designated as follows: TCS Pattern Xo, reflecting understanding, friendliness, and responsiveness *vs* aloofness and egocentrism on the part of the teacher; TCS Pattern Yo, reflecting responsible, businesslike, systematic *vs* evading, unplanned, slipshod teacher behaviour; and TCS Pattern Zo, reflecting stimulating, imaginative, original *vs* dull, routine teacher behaviour. Practical experience as well as the empirical data indicate that these are three of the principal areas involved in interpersonal relations, and that they might well be given basic consideration in the theory of teacher behaviour and also in teacher personnel procedures (pp. 102-103).

The simulation materials developed by Dale Bolton for the training of educational administrators in teacher selection are based on a sophisticated model of decision making. The following statements (Bolton, 1970) will give a general idea of the approach:

If all information about the teacher applicants is collected and then a single final selection decision is made, the data might take the form of multivariate information collected to predict various behavior which have varying utility in

relation to some institutional goal.

The tasks of the person who selects teachers include: (a) collecting reliable information, (b) using this information to predict consequent behaviors of the teacher, and (c) relating these behaviors to the operation of the organization so that a prediction of the total utility of the individual to the organization can be obtained.

These tasks are necessary to ascertain the relative merit of each applicant for a specific assignment. In addition, of course, the decision maker must determine how many (if any) of the applicants should be hired at a particular time. This decision depends on the quota to be filled at the time, the quality of the applicants being considered, the probability that additional people will apply, and the probable quality of additional applicants. The number of additional applicants and their quality are related to the time of the year (p. 3).

All decisions involve a consideration of the state of nature and a choice among alternatives. The choice is made on the basis of predicting (attaching a probability to) the consequences of the various alternatives and then assigning a value to the consequences predicted. It is the combination of the probable occurrence of an event and the value of the event that provides a utility measure for an alternative. The selection materials provided include the state of nature (the description of the situation), various alternatives (the applicants), a requirement to predict consequences (on the Teacher Evaluation Instrument), a value system (the explicit criteria for selection), and a choice (rank ordering of applicants) (p. 8).

In general, the decision-making process is based on two broad categories of data inputs: (1) descriptive information and (2) predicted outcomes. Each of these two categories operates within two domains -- the personal and the situational (or organizational). The relationships between these four aspects may be represented by a four-cell matrix as in Figure 1.

		DATA INPUT	
		DESCRIPTIVE INFORMATION	PREDICTED OUTCOMES
DOMAIN	PERSONAL		
	SITUATIONAL		

Figure 1

Relation of Information and Predictions to Personal and Situational Variables in the Selection of Applicants.

Extrapolating from *teacher selection* to the *selection of student applicants*, it becomes evident that there are many fewer variables to guide decision-making in the latter case. For instance, in the situational domain the variables are either common (such as, the demands of the B.Ed. program) or unknown (such as, the demands of particular positions in specific schools where graduates will ultimately find employment); being a constant in either case, this entire set of variables provides little help in discriminating between subjects. Likewise, the information category of the personal domain turns out to be of marginal use because all applicants must pass an initial screening of their academic admissibility. Once their scholastic eligibility has been established, the information category can provide very little data for justifying the admission of some and the exclusion of others. Figure 2 indicates that only the prediction of personal behaviour remains as a significant basis for discriminating between applicants.

		DATA INPUT	
		DESCRIPTIVE INFORMATION	PREDICTED OUTCOMES
DOMAIN	PERSONAL	Some basis for discrimination	Considerable basis for discrimination
	SITUATIONAL	Little basis for discrimination	Little basis for discrimination

Figure 2

The Utility of Different Variables in the Selection of Student Applicants to a Faculty of Education

For the prediction of consequent teacher variables, Bolton devised a machine-scorable teacher evaluation instrument which requires the rater to predict how the applicant will be evaluated at the end of one year of teaching in a pre-specified teaching post. Forty-nine bi-polar behavioural dimensions (e.g., stimulating-dull) are organized under nine "teaching acts" or professional contexts: planning and organizing classwork, classroom management, creating a motivational environment, instruction, evaluation, guidance and counselling, out-of-classroom professional activities, relations with staff and parents, and school-community relations. Scoring may be done in two ways. One way is to tally the scores of the scales in each of the separate role activities, thus obtaining nine professional role ratings. Another way, however, is to assign each of the 49 items into one of five major dimensions derived from Ryans' study. For this scoring approach, Bolton provides five templates representing organization, sociability, originality, empathy, and buoyancy. By adding the scores showing in the windows of each template, one can obtain summary ratings for the five generic factors, called important attributes.

The influence of Bolton's work on the Queen's interview rating instrument was very marked, for he had operationalized Ryans' empirical findings into an attractive and useful format for the real world of decision-making. However, because his materials were intended for use with certificated teachers (not applicants for teacher training) and were designed specifically for elementary teachers (not secondary school teachers), some considerable modification, as well as simplification, of his materials were required.

The final version of the Queen's Interview Assessment form (Figure 3) utilizes 30 bi-polar scales. Bolton's professional role contexts have been removed as a sharp focussing device and appear at the top as a general statement of instructions applying to the entire instrument. Because Queen's lacked matching scoring facilities, the instrument was redesigned for keypunch coding. Reducing turnaround time to an absolute minimum represented a major concern in all these plans since the prime function of the interview assessment forms was to serve pressing administrative purposes rather than research purposes alone.

It was hoped that, by requiring observers to make decisions about 30 discrete dimensions on the basis of 210 choice-points, some of the stereotyping that frequently characterizes personal interviews, particularly hurried ones, might be counteracted. Furthermore, the 30 items were arranged randomly so as to mask the six important attributes which they exemplified; namely empathy, organization, buoyancy, originality, leadership, and professional impression. The first four of these terms are the exact names Bolton assigned to his clusters of dimension. The other two attributes, leadership and professional impression, were original in nomenclature and in many of the component items. Figure 4 illustrates how the 30 items were classified into six attributes and indicates as well their source of origin. Given its bi-polar construction, each item consists of two terms, yielding a total of 60 terms for the 30 items. Of these 60 terms, 21 came from Ryans, 20 came from Bolton, and 19 came from Peruniak that were used to develop the final instrument presented in Appendix 2.

Two other features of the instrument should be mentioned. (1) For purposes of checking the observers' consistency and the instrument's

INTERVIEW ASSESSMENT

APPLICANT:

Student Number

1

Sex

8

Date

9

DAY MO. YEAR

Initial

15

Surname

16

INTERVIEWER:

No.

36

Status

39

(Code: T Teacher, A Administrator, F - Faculty, S Student)

Signature

PERSONAL DIMENSIONS

Indicate the rating for each dimension by marking an 'X' in the appropriate box on each line.

Since these dimensions should be considered in some professional context, the interviewer should estimate the candidate's potential in the light of these common professional roles:

As a manager of classroom instruction

As a member of staff & the school community

As a guide & counsellor of students

As a member of professional associations

		1	2	3	4	5	6	7	
Alone	40								Responsive
Dull	41								Stimulating
Disorganized	42								Systematic
Fuzzy	43								Precise communication
Irrresolute	44								Authoritative
Rigid	45								Adaptable
Autocratic	46								Democratic
Aimless	47								Purposeful
Pessimistic	48								Cheerful
Evading	49								Responsible
Unimpressive	50								Personal magnetism
Suspicious	51								Trusting
Easily upset	52								Self-possessed
Inflexible	53								Open-minded
Uninspiring	54								Challenging
Critical	55								Kindly
Humourless	56								Sense of humour
Distracting	57								Pleasant voice
Lethargic	58								Vigorous
Inconsistent	59								Consistent
Hindering	60								Helpful
Unimaginative	61								Resourceful
Expressionless	62								Expressive
Antagonistic	63								Co-operative
Dependent	64								Self-reliant
Stereotyped	65								Original
Irrational	66								Rational
Apathetic	67								Alert
Unintelligible	68								Fluent
Retiring	69								Forceful

ORDER OF IMPORTANT ATTRIBUTES (Rank 1 to 6)

Empathy	70	
Organization	71	
Leadership	72	
Buoyancy	73	
Originality	74	
Professional Impression	75	

(1 is the attribute which is most evident, and 6 is the attribute which is least evident.)

OVERALL IMPRESSION (MARK WITH 'X')

		1	2	3	4	5	6	7	
Unsuitable	76								Outstanding

COMMENTS:

Figure 3

Queen's Interview Assessment Form

ATTRIBUTES .	ITEMS		SOURCES	
			Left Term	: Right Term
PATHY	7	autocratic : democratic	(R) :	(R)
	12	suspicious : trusting	P :	P
	16	critical : kindly	B :	(R)
	21	hindering : helpful	B :	B
	24	antagonistic : co-operative	B :	B
ORGANIZATION	3	disorganized : systematic	(R) :	(R)
	8	aimless : purposeful	B :	B
	10	evading : responsible	(R) :	(R)
	20	inconsistent : consistent	B :	B
	27	irrational : rational	B :	B
LEADERSHIP	5	irresolute : authoritative	P :	P
	15	uninspiring : challenging	B :	P
	19	lethargic : vigorous	P :	P
	25	dependent : self-reliant	P :	P
	30	retiring : forceful	P :	P
EMOTIONALITY	1	aloof : responsive	(R) :	(R)
	9	pessimistic : cheerful	(R) :	P
	17	humourless : sense of humour	B :	B
	23	expressionless : expressive	P :	B
	28	apathetic : alert	(R) :	(R)
ORIGINALITY	2	dull : stimulating	(R) :	(R)
	6	rigid : adaptable	(R) :	(R)
	14	inflexible : open-minded	(R) :	B
	22	unimaginative : resourceful	B :	B
	26	stereotyped : original	(R) :	(R)
PROFESSIONAL IMPRESSION	4	fuzzy : precise comm.	B :	B
	11	unimpressive : personal magnetism	(R) :	P
	13	easily upset : self-possessed	P :	P
	18	distracting : pleasant voice	P :	(R)
	29	unintelligible : fluent	P :	P

R = Ryans

B = Bolton

P = Peruniak

Figure 4

Classification of Thirty Queen's Interview
Assessment Form Items by Attribute
and Source of Bi-polar Adjectives

content validity, a second section required raters to rank-order the important attributes displayed by each interviewee; results on this section could then be compared with the rank-ordered mean scores of the 30 items when grouped into attributes. (2) The final section of the instrument, titled Overall Impression, forced observers to make a global evaluative judgment in the form of a single score on a seven-point scale ranging from unsuitable to outstanding.

These then were the principal characteristics of the assessment instrument developed at the Faculty of Education, Queen's University, for interviewing purposes during 1971.

CHAPTER V

INTERVIEW PROCEDURES

The interviewing of academically qualified applicants for the new academic year was scheduled to begin in March, 1971. At that time, the faculty numbered 40, and the enrolment stood at 331. The projected enrolment figure for September, 1971, had been set at 650. Because interviews are conducted during the academic year preceding the year of actual attendance, it will be seen that, as the student enrolment was being doubled, the faculty complement available for interview purposes remained constant; new staff members would not be joining the Faculty until after the completion of the admissions phase.

Estimating that about 1,000 interviews would need to be conducted in order to recruit 650 candidates, the Faculty decided to stretch and augment staff resources in the following ways: (1) interviews would be limited to 30 minutes; (2) interviewing panels would consist of no more than three members; (3) outside personnel would be invited to assist faculty with the interviewing. Thus, a typical interviewing panel consisted of one faculty member (the chairman), one candidate-in-training, and one practicing teacher or administrator from the associate schools.

Training of observers could only be described as minimal. Less than half of the student and faculty interviewers attended a practice session, which featured the presentation of some interviewing guidelines, the rating of an applicant in a videotaped interview, and follow-up discussion. None of the assisting school personnel could be included in this orientation exercise, nor was any training provided to the 800 associate teachers who rated candidates during the subsequent practicum phase.

However, all interviewing panels were provided with a nine-page set of printed instructions, which was to be studied prior to the first interview (Appendix 3). These guidelines were not distributed to the associate

teachers in the follow-up stage (the practicum); they received only one page of general information about the project (Appendix 4).

Understandably, many of the individuals connected with this project entertained serious reservations about the usefulness of the enterprise. Some of the most fundamental principles governing indirect measurement appeared to be violated:

- (i) observers, by and large, were untrained;
- (ii) the procedures for observation, in the main, were nonsystematized;
- (iii) much of the assessment was based on inferred rather than observed categories;
- (iv) observational recording did not proceed concurrently with the events but was based on recall of past behaviour after the interview was over.

Therefore, there were grounds for fearing that raters' assessments might tend to be at considerable variance with one another, that they might reflect biases stemming from lack of criterion specificity, and that they might be contaminated by the unpredictable, non-directed nature of the interviewing transaction.

Nevertheless, once the policy decision had been made, reservations about the lack of that degree of methodological rigour demanded by fundamental research were not allowed to impede the arrangements. It was hoped that in the different world of professional practice even a modest degree of systematization might yield some useful dividends.

The main interviewing thrust was planned for March 18 and 19. All classes were cancelled for those two days in order to release faculty and students for that purpose. Interviewing took place at five centres: Kingston, Ottawa, Montreal, Toronto, and Kitchener. No difficulty was experienced in securing a slate of 40 enthusiastic students to become panel members. Likewise, the directs of education for the surrounding counties proved highly cooperative in releasing a total of 40 administrators and teachers to assist with the interviewing.

Approximately 465 interviews were conducted in that first round, leaving several hundred still to be held in the coming weeks and months. In subsequent interviews it was rarely possible to involve school personnel. Students also became unavailable after the term ended in early May. The registrar, who was responsible for scheduling interview times and identifying faculty interviewers, found the staff increasingly scarce. As the time-lag caused by the interview phase threatened to jeopardize the attainment of the Faculty's target figure, an administrative decision was made to grant admissions to the backlog of applicants in order to meet the quota and to postpone their personal interviews until the fall registration. The monthly distribution of interviewing was as follows: March, 465; April, 138; May, 95; June, 34; July, 2; August, 3; September, 99; October, 4. Over 840 applicants were interviewed to generate a class of 640.

Both the student members and the school members of the panels generally found the interviewing very stimulating and satisfying. Practicing teachers and administrators were also deeply impressed by the excellent calibre of young men and women seeking admission to the profession. Later, several directors of education endorsed with enthusiasm this model of faculty-school collaboration and urged that it be continued and expanded.

Faculty members, on the other hand, tended to be more polarized about the issue. Some genuinely questioned the predictive value of such interviews; a number doubted their own personal competence for this serious responsibility; a few resented the heavy demands of time. It was not altogether surprising, therefore, that the Faculty Board eventually eliminated the universal interview as an element in the admissions procedure for the following year.

If a hypothetical commentary may be permitted at this point, it is the conjecture that perhaps very few faculty would remain committed to heavy annual involvement in interview procedures -- unless selection was directly related to the composition of their own classes. Screening applicants for the good of the profession or for the ultimate benefit

of school children or even for the faculty of education to seem like an altruistic objective much too abstract and personally remote for most mortals.

There were two periods of practicum experience: November 1-12, 1971, and February 7-March 3, 1972. Associate teachers were requested by means of a one-page memorandum to complete an assessment instrument for each candidate whose practice teaching they observed. The level of returns in this follow-up phase would have to be characterized as good on the part of associates and fair-to-poor on the part of faculty.

Initially, there was some concern on the part of associates. However, the faculty in their visits to the schools were soon able to allay much of the anxiety and to correct most misperceptions. Nevertheless, a small number of associates sent in letters indicating that they were unable to participate in the study. Their grounds for refusing included violation of professional ethics, vagueness of the criteria, and their inability to assess personal qualities. A smaller number wrote letters of support commending the Faculty for undertaking a vital piece of research and for developing sophisticated instrumentation.

Another group with very strong feelings about the project were the actual subjects of the treatments -- the candidates. They communicated in no uncertain manner their sense of betrayal at not being informed about the personal assessment instrument prior to the first practicum. The explanation that it was judged unwise to burden them with additional anxiety for their first round of practice teaching and that, consequently, the announcement was planned for their return from the schools, received short shrift.

The negative reactions of some of the associates and of many of the candidates illustrate again the need to maintain desirable personal contact without jeopardizing the representativeness of the situation for research purposes. A recognition of the aversive side effects of certain research procedures, however, is only a first step towards a solution -- not the solution itself. The remaining part of the task poses some critical dilemmas; for instance, how can real freedom of choice be extended to the subjects and still ensure a representative sampling of the population?

On December 3, 1971, during the Professional Issues period, which the entire student body is expected to attend, a detailed explanation of the interview project was presented to the candidates. The assessment form was then distributed through the audience, and the students were requested to do a self-report.

After the venting of some initial hostility and skepticism, the overwhelming majority of students complied with the request; 344 returns came in, of which 14 were spoiled and 330 were usable. Since the absentee rate that morning approached 50%, it is obvious that the return rate from those in attendance was surprisingly high. Although a small percentage continued to find the whole thing repugnant, apparently most of the candidates were sufficiently persuaded of the institutional value of this study that they exposed themselves to the embarrassment of self-evaluation on highly personal dimensions.

CHAPTER VI

ANALYSIS OF THE DATA

The main purpose of this report is to describe the long-term project that initially was concerned with the development of an interview-rating instrument. Some six years later the focus was upon a comparison of interview results with ratings of teacher effectiveness. An initial overview of the development stages and analysis of the associated data is presented to, in part, demonstrate the basis for claims made about the interview results and to provide documentation for any in-depth follow-up that may result.

The data presented are based on returns from the following groups of raters and assessment contexts:

Group Identity	No. of Returns	Assessment Context
1. Faculty	933	Pre-Admission Interviews; Student Teaching
2. Students	467	Pre-Admission Interviews
3. Teachers	346	Pre-Admission Interviews
4. Administrators	254	Pre-Admission Interviews
5. Associates 1	660	Student Teaching (November)
6. Associates 2	778	Student Teaching (February)

It may be noted that assessments by groups 2, 3, and 4 were derived entirely from the pre-admission interviews. By contrast, assessments by groups 5 and 6 were derived from the supervision of students during practice teaching over periods of 2-4 weeks; we may properly regard these returns as post-admission data. The 933 returns from group 1 represent a mix of pre- and post-admission data; in certain respects it might have been advantageous to have kept the two Faculty categories separate.

Data from two other groups will be included incidentally in this report, namely: 7 - Student Self-Ratings (325 returns) and 8 - Composite (600 returns), representing a bolstered school educators' group attained by combining returns from 3 and 4.

An heuristic approach to the issue of inter-rater agreement, rather than a statistical one, will be presented. Evidence of stability and consistency of ratings will be sought in each of the three main sections of the instrument: (1) in the first section, "Personal Dimensions", the mean scores and standard deviations for the 30 items will be examined; (2) in the second section, "Order of Important Attributes", rank-order distributions will be compared; and (3) in the final section "Overall Impression", the global ratings will be analyzed.

Table 1 presents the mean scores assigned to each of the 30 items by each of the six rating groups. Examination of the highest and lowest means shows very little spread for any of the group; for example, the administrators scored a low of 4.62 (on item 15) and a high of 5.36 (on items 24 and 28), representing a difference of only .74. The greatest difference was recorded by Associate 1, with a low mean of 5.54 assigned to item 30 and a high mean of 6.07 assigned to item 24, yielding a difference of 1.53, which is still quite minimal considering that it covers a rating scale of seven intervals and a range of 30 disparate items.

Table 2 presents the standard deviations for the scores assigned to each of the 30 items by the six rating groups. Once again the general impression is that of relatively little variation.

The section titled, "Order of Importance Attributes", represented the second rating task for interviewers, following the first task which was scoring the 30 items under "Personal Dimensions". On the basis of their observation of the student, raters were asked to rank-order the six specified attributes; namely, empathy, organization, leadership, buoyancy, originality, and professional impression. Raters assigned 1 to the most prominent attribute and 6 to the least evident.

It will be recalled that the purpose of this section was to provide a check on internal consistency. Would the rank-ordering of the six

Table 1

Mean Scores on 30 Items, Assigned
by Six Groups

Item	Faculty (N=933)	Students (N=467)	Teachers (N=346)	Admin. (N=254)	Assoc. 1 (N=660)	Assoc. 2 (N=778)
1	5.58	5.50	5.71	5.16	5.36	5.30
2	5.20	4.96	5.28	4.92	4.69	4.80
3	5.35	5.23	5.27	4.94	5.10	5.07
4	5.29	5.05	5.16	4.91	4.69	4.77
5	5.09	4.87	4.81	4.69	4.58	4.64
6	5.32	5.11	5.31	5.05	5.42	5.37
7	5.32	5.19	5.34	5.09	5.33	5.28
8	5.68	5.46	5.56	5.23	5.32	5.29
9	5.65	5.62	5.60	5.27	5.61	5.59
10	5.78	5.60	5.58	5.31	5.75	5.65
11	5.09	4.84	5.05	4.74	4.78	4.85
12	5.45	5.22	5.39	5.22	5.61	5.51
13	5.57	5.34	5.35	5.08	5.34	5.31
14	5.46	5.27	5.40	5.13	5.58	5.46
15	4.95	4.71	4.92	4.62	4.62	4.74
16	5.54	5.59	5.55	5.37	5.66	5.51
17	5.16	5.09	5.14	4.80	5.25	5.31
18	5.37	5.45	5.36	5.14	5.28	5.38
19	5.22	4.96	5.26	4.86	4.89	4.95
20	5.51	5.35	5.42	5.05	5.38	5.32
21	5.65	5.49	5.63	5.27	5.84	5.70
22	5.28	5.07	5.22	4.97	5.16	5.20
23	5.26	5.07	5.28	5.00	4.84	4.93
24	5.75	5.60	5.81	5.36	6.07	5.88
25	5.67	5.33	5.40	5.11	5.23	5.35
26	4.96	4.73	4.92	4.70	4.73	4.85
27	5.59	5.47	5.65	5.20	5.63	5.51
28	5.60	5.47	5.65	5.36	5.44	5.37
29	5.42	5.39	5.55	5.19	5.36	5.34
30	4.96	4.68	4.94	4.69	4.54	4.65
High Mean	5.78	5.62	5.81	5.36	6.07	5.88
Low Mean	4.95	4.68	4.81	4.62	4.54	4.64
Mean of Means	5.39	5.22	5.35	5.05	5.24	5.23

Table 2

Standard Deviations on 30 Items Scored
by Six Groups

Item	Faculty (N=933)	Students (N=467)	Teachers (N=346)	Admin. (N=254)	Assoc. 1 (N=660)	Assoc. 2 (N=778)
1	1.14	1.15	1.15	1.06	1.16	1.20
2	1.33	1.34	1.26	1.33	1.24	1.27
3	1.13	1.15	1.19	1.11	1.19	1.26
4	1.30	1.34	1.50	1.23	1.23	1.31
5	1.23	1.23	1.31	1.19	1.25	1.19
6	1.21	1.34	1.13	1.18	1.19	1.20
7	1.15	1.11	1.17	1.12	1.14	1.15
8	1.20	1.13	1.19	1.23	1.18	1.19
9	1.20	1.23	1.19	1.23	1.27	1.24
10	1.12	1.11	1.18	1.21	1.18	1.26
11	1.42	1.48	1.38	1.55	1.43	1.33
12	1.03	1.15	1.16	1.09	1.04	1.09
13	1.14	1.25	1.38	1.20	1.26	1.23
14	1.15	1.20	1.13	1.20	1.13	1.14
15	1.35	1.38	1.33	1.37	1.36	1.28
16	1.09	1.15	1.18	1.10	1.04	1.13
17	1.26	1.31	1.21	1.31	1.26	1.17
18	1.24	1.25	1.29	1.14	1.33	1.15
19	1.27	1.34	1.29	1.30	1.34	1.30
20	1.05	0.99	1.13	1.09	1.01	1.07
21	0.97	1.09	0.94	0.96	0.92	1.05
22	1.32	1.40	1.34	1.31	1.34	1.31
23	1.32	1.39	1.37	1.32	1.32	1.22
24	0.96	1.07	0.94	1.04	0.96	1.04
25	1.23	1.29	1.37	1.28	1.40	1.33
26	1.21	1.32	1.13	1.21	1.27	1.18
27	0.97	0.94	1.02	1.04	0.98	1.00
28	1.15	1.14	1.16	1.26	1.12	1.20
29	1.22	1.14	1.21	1.13	1.14	1.09
30	1.39	1.52	1.44	1.33	1.38	1.33

attributes in section two be congruent with the rank-ordering derived from the scores on the 30 items of section one? For the purpose of rank-ordering the data in section one, the mean scores of the five component items comprising each attribute cluster were summed. The following table identifies the component items within each attribute:

	Empathy	Organi- zation	Leader- ship	Buoy- ancy	Origi- nality	Professional Impression
	7	3	5	1	2	4
	12	8	15	9	6	11
Items	16	10	19	17	14	13
	21	20	25	23	22	18
	24	27	30	28	26	29

A summary comparison of the rank-ordering results from the two sections appears in Table 3. The two patterns are far from congruent. Although raters consistently place empathy and organization in the first two ranks in both sections, the ordering of the remaining four attributes shows no relationship between sections. These disappointing results suggest to the investigator that (1) the attributes are low in content validity, and/or (2) the rank-ordering task section two is far too difficult, given the high number of variables (30 items) to be mentally processed.

If Table 3 is discouraging regarding comparison *between* sections, it yet further corroborates the existence of stability among raters *within* the two sections viewed independently. We have already noted this phenomenon in the earlier discussion on mean scores assigned to the 30 items. It seems significant that a similar stability characterizes the second section, where raters were assessing more comprehensive or global dimensions; for example, each of the six groups without exception rank-ordered Originality as 6.

Table 4 indicates the high degree of consistency between mean scores on the overall impression and the ratings on the 30 personal dimensions

TABLE 3

Comparison of Rank-Ordering of Attributes in Two Sections

	Faculty	Students	Teachers	Adminis- tration	Assoc. 1	Assoc. 2
Empathy	3.27 ...1	3.22 ...1	3.23 ...1	2.77 ...1	2.97 ...2	3.12 ...2
Organization	3.56 ...2	3.39 ...2	3.46 ...2	3.19 ...2	2.92 ...1	3.02 ...1
Leadership	3.84 ...3	3.41 ...3	3.77 ...4	3.63 ...3	3.81 ...4	3.78 ...4
Buoyancy	4.10 ...5	4.09 ...5	3.77 ...3	3.78 ...4	4.03 ...5	4.03 ...5
Originality	4.84 ...6	4.78 ...6	4.63 ...6	4.48 ...6	4.41 ...6	4.41 ...6
Professional Impression	3.97 ...4	3.96 ...4	3.90 ...5	3.86 ...5	3.53 ...3	3.37 ...3

Rank Orders in "Order of Important Attributes" Section,
Showing the Means of Rank Orders Assigned to Each Attribute

	Faculty	Students	Teachers	Adminis- tration	Assoc. 1	Assoc. 2
Empathy	27.71 ...2	27.09 ...2	27.72 ...1	26.31 ...1	28.51 ...1	27.88 ...1
Organization	27.91 ...1	27.11 ...1	27.48 ...2	25.73 ...2	27.18 ...2	26.84 ...2
Leadership	25.89 ...6	24.55 ...6	25.33 ...6	23.97 ...6	23.86 ...6	24.33 ...6
Buoyancy	27.25 ...3	26.75 ...3	27.38 ...3	25.59 ...3	26.50 ...3	26.70 ...3
Originality	26.22 ...5	25.14 ...5	26.13 ...5	24.77 ...5	25.58 ...4	25.68 ...4
Professional Impression	26.74 ...4	26.07 ...4	26.47 ...4	25.06 ...4	25.45 ...5	25.65 ...5

Rank Orders Derived from 30 Items in "Personal Dimensions" Section,
Showing the Sums of Mean Scores on the Five Component Items Within Each Attribute

when reduced to the means of mean scores.

Table 4
Comparison of Overall Impression Ratings
and Personal Dimensions Ratings

	Faculty	Students	Teachers	Adminis- tration	Assoc. 1	Assoc. 2
Mean Scores on Overall Impression	5.23	5.10	5.27	4.80	4.98	5.10
Mean of Mean Scores on 30 Items	5.39	5.22	5.35	5.05	5.24	5.23

In light of the above comparison, it could be argued that in the interest of economy and simplicity the single overall impression rating could suffice as the admissions index, rendering a longer assessment instrument unnecessary. The investigator's view, however, is that a more valid overall impression score is produced when observers have been required to identify and to rate a number of discrete dimensions, as a preliminary exercise in consciousness-raising, before attempting a molar estimate.

Support for construct validity is provided in the following section, where the principal factors are seen to be stable across five of the six groups. It is to be hoped that, in due course, statistical verification may be provided, but it is anticipated that this statistic will merely corroborate what is already apparent: the instrument yields consistent and similar sets of scores even when used by relatively inexperienced and untrained personnel utilizing a global procedure based largely upon independent judgments.

Discussion of validity in this section will focus mainly on content validity, with only brief reference to criterion-related validity which

is the focus in the later set of results obtained six years after the students' graduation.

Content validity provides evidence about the degree to which the measures of an instrument represent those distinct qualities, abilities, or attitudes which they purport to estimate. Such corroboration is particularly vital in the case of inferred mental or personality characteristics for which constructs are postulated. If it can be established that the assessment scores do, in fact, correspond to the phenomenon claimed, then a sound base is provided for subsequently demonstrating the instrument's predictive utility with reference to certain terminal or output variables. One powerful, but indirect, approach to content validation is through factor analysis, where it can be demonstrated that items cluster in patterns similar to those defined by previous researchers, such as Ryans and Bolton in this case.

As the first step, collected data were translated into a 30 x 30 item inter-correlation matrix for each group of raters. It seems significant that in none of the eight matrices does a single negative correlation appear; in short, most of the items appear to be positively correlated. One pair of items which consistently had low inter-correlations, such that they may be considered essentially uncorrelated, were irresolute-authoritative and autocratic-democratic. The single lowest correlation emerged in the teachers' matrix, where retiring-forceful correlated with critical-kindly ($r = .005$).

Each inter-correlation matrix was then analyzed using the principal component method, a data-reduction procedure which attempts to explain the variability by extracting variance-saturated components or factors of descending magnitude until the total common variance (30 in this case) has been accounted for. From the accompanying table of eigenvalues (Table 5), it may be seen that the first factor in every case accounts for a high proportion of the total variance, averaging out to roughly 50% across the six groups. The fourth, fifth, and sixth factors given some indication of the rapid loss of explanatory power in the remaining 24 factors which are not shown here.

Table 5

Eigenvalues for Six Factors by Rating Groups

Factors	Faculty	Students	Teachers	Adminis- tration	Assoc. 1	Assoc. 2
1	17.197	14.109	14.224	17.760	13.032	14.571
2	2.206	2.527	2.813	2.143	2.237	2.370
3	1.430	1.835	1.667	1.187	1.796	1.615
4	0.799	1.126	1.098	0.893	1.131	1.024
5	0.726	0.779	0.875	0.738	0.940	0.842
6	0.574	0.746	0.701	0.618	0.887	0.792
	22.93	21.12	22.38	23.34	20.02	21.22

The next step was to transform the principal components solution into a more preferred solution by rotating the reference axes in accordance with the Varimax Criterion technique. The general purpose of this procedure is to clarify the salient features of the factor so that the conceptual identity of the group of high-loading items may be ascertained and understood.

Tables 6 and 7 depict a six-factor Varimax solution for faculty and administrators, with item loadings grouped by the specified attributes. Several features deserve comment. (1) The factor structure is dissimilar in the two solutions, with the faculty pattern representative of all the other rater groups and the administrators' pattern providing the single maverick. (2) If one considers the items classified within each given attribute, many, but not all, load high on the same factor. Apparently, there is a measure of conceptual commonality between certain attributes and certain factors.

Table 6

Varimax Factor Loadings for 30 Items
Grouped by Attributes (Faculty: N = 933)

Attributes	Item No.	Factors						Commonalities
		I	II	III	IV	V	VI	
Empathy	7	.124	.574	.124	.654	.055	-0.027	.791
	12	.233	.730	.193	.184	.157	.014	.683
	16	.110	.783	.118	.255	.146	.108	.737
	21	.222	.692	.335	.214	.139	.237	.761
	24	.243	.711	.269	.152	.152	.180	.716
Organization	3	.263	.204	.788	.117	.127	.018	.762
	8	.415	.303	.659	.173	-0.043	.129	.746
	10	.348	.569	.548	.114	-0.008	.058	.761
	20	.237	.391	.565	.129	.110	.407	.723
	27	.164	.414	.580	.163	.222	.380	.754
Leadership	5	.545	.004	.574	.125	.087	.148	.664
	15	.704	.219	.377	.305	.174	.107	.819
	19	.792	.269	.246	.050	.115	.156	.790
	25	.564	.206	.437	.133	.002	.454	.717
	30	.800	.054	.282	.052	.130	.226	.790
Buoyancy	1	.581	.520	.264	.187	.034	-0.091	.722
	9	.605	.644	.120	.089	.050	-0.120	.820
	17	.594	.555	.023	.166	.182	-0.003	.722
	23	.724	.265	.225	.220	.281	.036	.773
	28	.607	.390	.345	.114	.211	.171	.726
Originality	2	.722	.270	.346	.288	.192	-0.079	.839
	6	.299	.462	.235	.652	.088	.006	.790
	14	.224	.571	.184	.592	.120	.176	.806
	22	.526	.263	.404	.430	.112	.243	.766
	26	.488	.187	.316	.516	.153	.316	.762
Professional Expression	4	.347	.160	.712	.233	.375	-0.082	.854
	11	.731	.352	.241	.238	.211	-0.040	.819
	13	.557	.200	.384	.156	.172	.315	.651
	18	.427	.370	.117	.118	.672	.106	.809
	29	.391	.230	.460	.112	.620	.045	.816
Eigenvalues		7.438	5.721	4.709	2.420	1.539	1.105	22.932
Relative Proportion of the Common Variance		32	57	78	88	95	100	

Tables 8, 9, 10, 11, 12, and 13 show the items loading above .65, rank-ordered for each of the six factors. It will be seen that not a single high-loading item appears under Factor VI for any of the six rating groups. Similarly, few high-loading items fall out in Factors IV and V, the largest clusters containing only two items. In addition, the high-loading items in Factors IV and V display little commonality between rater groups; we seem to be dealing with elements which are disparate, residual, and group-specific rather than consistent and generic in nature.

Because of the foregoing considerations, the investigator determined to focus attention on the first three factors from the six-factor varimax solution. The first three factors appear to provide a more solid basis for a discussion of item validation (and possibly eventual item reduction). Further analyses using three principal components followed by a varimax rotation would be a desirable extension for exploration of the underlying structure. However, that activity was not undertaken for purposes of this presentation and must be relegated to a follow-up study.

Table 14 presents frequency distribution of the high-loading items for five of the six groups of raters, the administrators being excluded. On examining the items loading highly on Factor III, one observes dimensions such as (3) disorganized-systematic, (8) aimless-purposeful, (30) inconsistent-consistent, (4) fuzzy-precise communication, (27) irrational-rational. It is apparent that Factor III represents the concept of organization. (How the two items, (6) rigid-adaptable and (7) autocratic-democratic, came to be included in Factor III by the students' group is not clear.)

An analysis of Factor II's high-load items reveals dimensions such as (16) critical-kindly, (12) suspicious-trusting, (24) antagonistic-cooperative, (7) autocratic-democratic, (14) inflexible-open-minded, (6) rigid-adaptable, and (21) hindering-helpful. It seems clear that Factor II represents the concept of empathy.

The salient items in Factor I include (2) dull-stimulating, (11) unimpressive-personal magnetism, (19) lethargic-vigorous,

Table 7

Varimax Factor Loadings for 30 Items
Grouped by Attributes (Admin.: N = 254)

Attributes	Item No.	Factors						Commonalities
		I	II	III	IV	V	VI	
Empathy	7	.682	.009	.320	.080	.331	-0.083	.691
	12	.765	.316	.264	.049	-0.112	.027	.770
	16	.789	.176	.267	.107	.017	.137	.755
	21	.680	.365	.177	.125	.273	.039	.719
	24	.781	.232	.178	.240	.113	.049	.769
Organization	3	.281	.710	.171	.162	.196	.318	.778
	8	.271	.670	.279	.187	.334	.091	.755
	10	.609	.570	.207	.226	.031	.067	.796
	20	.435	.500	.170	.293	.358	.223	.733
	27	.595	.357	.140	.361	.327	.062	.742
Leadership	5	.095	.807	.255	.113	.010	.087	.747
	15	.414	.421	.531	.172	.375	.165	.828
	19	.183	.560	.545	.165	.303	-0.047	.766
	25	.357	.601	.302	.178	.337	-0.223	.774
	30	.066	.702	.388	.239	.185	-0.254	.803
Buoyancy	1	.396	.305	.630	.196	.020	-0.251	.748
	9	.411	.235	.656	.170	.128	.038	.700
	17	.349	.237	.742	.127	.182	.190	.814
	23	.176	.373	.594	.452	.300	.009	.817
	28	.279	.457	.500	.340	.320	-0.072	.761
Originality	2	.293	.429	.696	.158	.186	-0.089	.821
	6	.734	.149	.322	.186	.282	-0.167	.806
	14	.772	.059	.302	.082	.305	.038	.792
	22	.378	.425	.369	.146	.586	.016	.824
	26	.357	.349	.398	.153	.605	-0.002	.798
Professional Expression	4	.264	.547	.320	.549	.010	.216	.820
	11	.417	.332	.703	.085	.201	.087	.834
	13	.356	.628	.311	.170	.123	-0.164	.689
	18	.302	.231	.525	.438	.043	.403	.777
	29	.265	.318	.254	.801	.173	-0.061	.911
Eigenvalues		6.739	5.993	5.379	2.265	2.221	0.741	23.338
Relative Proportion of the Common Variance		29	55	78	87	97	100	

Table 8

Varimax Factor Loadings for 30 Items (Faculty: N = 933),
with High-Loading Items (Higher than .65) Rank-Ordered by Clusters

Item No.	I	II	III	IV	V	VI	Common- alities
30	.800	.054	.284	.052	.130	.216	.790
19	.792	.269	.236	.050	.105	.146	.790
11	.731	.352	.241	.238	.211	-0.040	.819
23	.724	.265	.225	.220	.281	.036	.773
2	.722	.270	.346	.288	.192	-0.079	.839
15	.704	.229	.367	.295	.174	.097	.809
16	.110	.783	.118	.255	.146	.108	.737
12	.233	.730	.193	.184	.157	.014	.683
14	.243	.711	.269	.152	.152	.180	.716
21	.222	.692	.335	.214	.139	.237	.761
3	.263	.204	.788	.117	.127	.018	.762
4	.347	.160	.712	.233	.375	-0.082	.854
8	.415	.303	.659	.173	-0.043	.129	.746
7	.125	.574	.124	.654	.055	-0.027	.791
6	.299	.462	.235	.652	.088	.006	.790
8	.427	.370	.117	.118	.672	.106	.809

1	.581	.520	.264	.187	.034	-0.091	.722
5	.545	.004	.574	.115	.077	.138	.664
9	.605	.644	.120	.089	.050	-0.120	.820
10	.348	.569	.548	.114	-0.008	.058	.761
13	.557	.200	.384	.156	.172	.315	.651
14	.224	.571	.184	.592	.120	.176	.806
17	.594	.555	.023	.166	.182	-0.003	.722
20	.237	.391	.565	.129	.110	.407	.723
22	.526	.263	.404	.430	.112	.243	.766
25	.564	.206	.427	.133	.002	.454	.767
26	.488	.187	.316	.516	.153	.316	.762
27	.164	.414	.580	.163	.222	.380	.754
28	.607	.390	.345	.114	.211	.171	.726
29	.391	.230	.460	.112	.620	.045	.816
Eigen- values	7.438	5.721	4.709	2.420	1.539	1.105	22.932
Cumulative Proportion of Common Variance	32	57	78	88	95	100	

Table 9

Varimax Factor Loadings for 30 Items (Students: N = 467),
with High-Loading Items (Higher than .65) Rank-Ordered by Clusters

Item No.	I	II	III	IV	V	VI	Common- alities
19	.761	.135	.119	.198	.292	.104	.748
2	.759	.146	.320	.243	.142	.143	.799
23	.737	.185	.276	.149	.181	.200	.749
17	.721	.344	.165	.012	.096	.143	.696
11	.716	.275	.252	.118	.183	.232	.753
9	.713	.431	.154	.070	-0.102	.107	.745
28	.654	.249	.149	.206	.238	.234	.666
16	.264	.754	.310	.044	.071	.068	.746
24	.217	.712	.271	.153	.137	.119	.683
12	.241	.708	.254	.138	.095	.134	.669
7	.171	.349	.744	.065	-0.003	.070	.714
14	.171	.393	.737	.098	.101	.117	.760
6	.267	.341	.735	.015	.050	.071	.736
3	.126	.061	.077	.764	.184	.345	.761
8	.384	.237	.113	.683	.167	.006	.711
13	.252	.248	.108	.142	.717	.167	.699
15	.324	.071	.234	.246	.681	.077	.695
<hr/>							
1	.622	.325	.322	.266	.104	.011	.678
4	.301	.026	.155	.474	.237	.597	.752
5	.406	-0.029	-0.061	.496	.453	.097	.630
10	.272	.369	.249	.525	.180	.103	.591
15	.637	.091	.359	.296	.326	.138	.756
18	.315	.451	.116	.073	.127	.571	.663
20	.035	.306	.220	.408	.456	.276	.594
21	.255	.566	.352	.232	.193	.203	.642
22	.401	.107	.577	.248	.313	.105	.675
26	.422	.046	.537	.121	.350	.153	.629
27	.004	.362	.043	.397	.446	.404	.653
29	.287	.170	.141	.200	.146	.749	.753
30	.649	-0.063	-0.003	.191	.543	.132	.773
Eigen- values	6.433	3.657	3.438	2.813	2.740	2.039	21.121
Cumulative Proportion of Common Variance	30	48	64	77	90	100	

Table 10

Varimax Factor Loadings for 30 Items (Teachers: N = 346),
with High-Loading Items (Higher than .65) Rank-Ordered by Clusters

Item No.	I	II	III	IV	V	VI	Common- alities
19	<u>.837</u>	.086	.094	.158	.158	.208	.810
30	<u>.828</u>	-0.057	.299	.027	.118	.067	.797
2	<u>.750</u>	.244	.268	.223	.136	.047	.764
23	<u>.687</u>	.313	.159	.541	.007	.028	.799
11	<u>.667</u>	.355	.246	.308	.234	.146	.802
15	<u>.667</u>	.278	.367	.201	.111	.167	.737
5	<u>.655</u>	-0.059	.463	.053	.241	.050	.710
7	.128	<u>.831</u>	.140	.076	.089	.029	.740
6	.221	<u>.788</u>	.134	.023	.093	.004	.697
14	.096	<u>.787</u>	.229	.093	.015	.079	.697
16	-0.064	<u>.669</u>	.136	.058	.153	.506	.753
12	.193	<u>.651</u>	.116	.260	.344	.109	.672
3	.271	.146	<u>.808</u>	.143	.028	.005	.770
27	.205	.263	<u>.703</u>	.257	.003	.161	.698
20	.288	.165	<u>.682</u>	.132	.034	.378	.737
8	.402	.267	<u>.658</u>	.053	.035	.104	.681
18	.219	.149	.142	<u>.804</u>	.165	.214	.810
29	.354	.148	.407	<u>.717</u>	.052	-0.085	.836
<hr/>							
1	.569	.443	.235	.201	.267	-0.090	.692
4	.405	.200	.609	.448	.122	-0.080	.797
9	.429	.416	.079	.195	.629	.111	.810
10	.258	.356	.593	.076	.190	.398	.746
13	.534	.191	.494	.140	.337	-0.162	.726
17	.530	.306	.087	.253	.537	.178	.767
21	.242	.447	.345	.140	.079	.593	.755
22	.527	.451	.460	.097	-0.151	.117	.739
24	.229	.616	.255	.282	-0.044	.361	.709
25	.646	.125	.468	.089	-0.046	-0.067	.667
26	.601	.438	.295	.124	-0.290	.052	.743
28	.635	.251	.376	.305	.110	.082	.719
Eigen- values	7.261	5.129	4.837	2.363	1.424	1.363	22.378
Cumulative Proportion of Common Variance	32	55	77	88	94	100	

Table 11

Varimax Factor Loadings for 30 Items (Administrators: N = 254),
with High-Loading Items (Higher than .65) Rank-Ordered by Clusters

Item No.	<u>I</u>	<u>II</u>	<u>III</u>	<u>IV</u>	<u>V</u>	<u>VI</u>	Common- alities
24	.781	.232	.178	.240	.113	.049	.769
16	.789	.176	.267	.107	.017	.137	.755
14	.772	.059	.302	.082	.305	.038	.792
12	.765	.316	.264	.049	-0.112	.027	.770
6	.734	.149	.322	.186	.282	-0.167	.806
7	.682	.009	.320	.080	.331	-0.083	.691
21	.680	.365	.177	.125	.273	.039	.719
5	.095	.807	.255	.113	.010	.087	.747
3	.281	.710	.171	.162	.196	.318	.778
30	.066	.702	.388	.239	.185	-0.254	.803
8	.271	.670	.279	.187	.334	.091	.755
17	.349	.237	.742	.127	.182	.190	.814
11	.417	.332	.703	.085	.201	.087	.834
2	.293	.429	.696	.158	.186	-0.089	.821
9	.411	.235	.656	.170	.128	.038	.700
29	.265	.318	.254	.801	.173	-0.061	.911

1	.396	.305	.630	.196	.020	-0.251	.748
4	.264	.547	.320	.549	.010	.216	.820
10	.609	.570	.207	.226	.031	.067	.796
13	.356	.628	.311	.170	.123	-0.164	.689
15	.414	.421	.501	.172	.375	.165	.828
18	.302	.231	.525	.438	.043	.403	.777
19	.183	.560	.545	.165	.303	-0.047	.766
20	.435	.500	.170	.293	.358	.223	.733
22	.378	.425	.369	.146	.586	.016	.824
23	.176	.373	.594	.452	.300	.009	.817
25	.357	.601	.302	.178	.337	-0.223	.774
26	.357	.349	.398	.153	.605	-0.002	.798
27	.595	.357	.140	.361	.327	.062	.742
28	.279	.457	.500	.340	.320	-0.072	.761
Eigen- values	6.739	5.993	5.379	2.265	2.221	.741	23.338
Cumulative Proportion of Common Variance	29	55	78	87	97	100	

Table 12

Varimax Factor Loadings for 30 Items (Assoc. 1: N = 660),
with High-Loading Items (Higher than .65) Rank-Ordered by Clusters

Item No.	I	II	III	IV	V	VI	Common- alities
11	<u>.776</u>	.228	.176	.172	.187	.100	.760
19	<u>.775</u>	.233	.272	.040	.019	.058	.733
2	<u>.772</u>	.167	.186	.052	.218	.253	.772
23	<u>.744</u>	.217	.198	.040	.223	.167	.718
30	<u>.709</u>	-0.021	.249	.313	.229	.034	.717
15	<u>.673</u>	.148	.299	.168	.214	.310	.735
16	.144	<u>.712</u>	.151	-0.077	.083	.166	.591
24	.138	<u>.712</u>	.348	.059	.052	-0.097	.662
3	.189	.087	<u>.744</u>	.050	.025	.111	.612
13	.189	.191	.189	<u>.818</u>	.107	.076	.794
18	.397	.256	-0.019	.132	<u>.654</u>	-0.012	.668
29	.180	.166	.375	.052	<u>.715</u>	.126	.730

1	.487	.577	.122	.102	.015	.112	.608
4	.274	.040	.550	.185	.439	.146	.627
5	.446	-0.063	.407	.480	.233	-0.072	.650
6	.193	.494	.070	.362	.205	.427	.642
7	.050	.617	-0.149	.062	.283	.371	.627
8	.383	.212	.645	.120	.054	.216	.672
9	.611	.527	.026	.164	-0.038	.005	.680
10	.321	.526	.459	.158	.023	.057	.619
12	.200	.600	.184	.045	.066	.005	.440
14	.103	.625	.171	.347	.176	.302	.672
17	.604	.500	-0.062	.130	.095	.061	.648
20	.097	.282	.643	.185	.143	.024	.558
21	.225	.599	.430	.082	.052	.034	.605
22	.416	.236	.459	.058	.041	.555	.752
25	.419	.158	.395	.489	-0.048	.225	.648
26	.503	.166	.299	.100	.057	.630	.780
27	.098	.448	.582	.157	.204	.040	.617
28	.589	.362	.421	.138	.037	.021	.675
Eigen- values	6.214	4.795	3.979	1.815	1.683	1.537	20.022
Cumulative Proportion of Common Variance	31	55	75	84	92	100	

Table 13

Varimax Factor Loadings for 30 Items (Assoc. 2: N = 778),
with High-Loading Items (Higher than .65) Rank-Ordered by Clusters

Item No.	I	II	III	IV	V	VI	Common- alities
2	.755	.249	.246	.003	.174	.255	.789
30	.744	-0.021	.215	.375	.139	.040	.761
11	.743	.314	.196	.106	.232	.109	.765
23	.738	.281	.226	.037	.174	.129	.722
19	.735	.213	.321	.147	-0.002	.089	.718
15	.660	.240	.346	.040	.153	.342	.755
16	.198	.751	.195	-0.072	.104	-0.009	.657
12	.176	.721	.170	.212	.071	.025	.630
7	.051	.717	.012	-0.052	.179	.293	.637
24	.115	.688	.429	.176	.010	.008	.702
14	.116	.679	.194	.287	.073	.365	.733
3	.235	.104	.800	.017	.105	.121	.732
20	.262	.294	.718	.161	.086	.036	.705
8	.363	.163	.695	.074	.162	.144	.693
13	.187	.345	.216	.678	.232	.109	.727
18	.309	.334	.103	.111	.718	-0.063	.750

1	.517	.534	.301	-0.067	.027	.035	.650
4	.293	.095	.639	.039	.392	.259	.725
5	.513	-0.118	.392	.415	.174	-0.021	.634
6	.163	.575	.217	.180	.176	.428	.651
9	.513	.612	.201	.150	.069	-0.113	.718
10	.314	.399	.626	.209	-0.098	.108	.715
17	.483	.559	.054	.065	.259	.011	.620
21	.203	.550	.582	.115	.053	.082	.705
22	.445	.230	.463	.100	-0.036	.546	.774
25	.385	.170	.445	.500	.020	.329	.734
26	.484	.169	.356	.159	.087	.561	.738
27	.134	.301	.638	.307	.186	.106	.655
28	.517	.277	.507	.255	.118	.098	.689
29	.223	.133	.408	.221	.636	.210	.731
Eigen- values	5.876	5.310	5.247	1.634	1.586	1.561	21.214
Cumulative Proportion of Common Variance	28	53	77	85	93	100	

Frequency Distributions of High Loading Items on
Three Principal Factors for Five Rater Groups

Factor I

Item No.	Faculty	Students	Teachers	Associates 1	Associates 2	Total
2	X	X	X	X	X	5
11	X	X	X	X	X	5
19	X	X	X	X	X	5
23	X	X	X	X	X	5
30	X		X	X	X	4
15	X		X	X	X	4
17		X				1
5			X			1
9		X				1
28		X				1

Factor II

Item No.	Faculty	Students	Teachers	Associates 1	Associates 2	Total
16	X	X	X	X	X	5
12	X	X	X		X	4
24	X	X		X	X	4
7			X		X	2
14			X		X	2
6			X			1
21	X					1

Factor III

Item No.	Faculty	Students	Teachers	Associates 1	Associates 2	Total
3	X		X	X	X	4
8	X		X		X	3
20			X		X	2
4	X					1
6		X				1
7		X				1
27			X			1

(23) expressionless-expressive, (30) retiring-forceful, (15) uninspiring-challenging, (17) humourless-sense of humour, (5) irresolute-authoritative, (9) pessimistic-cheerful, and (28) apathetic-alert. It seems that Factor I represents the concept of dynamic vitality, to which we will assign the name surgency, borrowed from Cattell (Ryans, p. 111).

The conclusion to which this argument leads is that the first three factors of the Varimax six-factor solution support the content validity of the *high*-loading items in Table 14 and that one may use them (and possibly their generic equivalents) with considerable confidence that they are measuring what they purport to represent. (No such claim can be made for the other items).

The results of this analysis based on the first three factors correspond closely with Ryans' work. His TCS Pattern X (kindly, understanding, democratic *vs.* aloof, restricted, egocentric teacher (behaviour) seems related to the Empathy of Factor II. His TCS Pattern Y (responsible, systematic *vs.* evading, unplanned teacher behaviour) appears to involve the same behaviour as the Organization of Factor III. Finally, Ryans' TCS Pattern Z (stimulating, imaginative *vs.* dull, routine teacher behaviour) would seem to represent the Surgency of Factor I.

It is particularly gratifying when content validation can be established not only by factor analysis of the data but can be corroborated as well by the findings of the original investigation. To the extent that the results of the present study replicate those of Ryans', it may be argued that it has served a cross-validation function, providing mutually-supporting evidence for the claim of validity.

Ideally, a study such as this ought to address the question of predictive validity. Do the interview scores serve to predict applicants' teaching effectiveness? For a number of reasons, that question cannot be answered from this data alone.

CHAPTER VII

RESULTS AND CONCLUSIONS

A faculty of education selection process is designed to select the "best" set of applicants to admit. Typically, "best" has been defined as those with the highest undergraduate grades, the highest test scores, and the best performance on an interview. There are many reasons why best has been defined by these measures, but the basic argument is that these will predict academic performance in faculty of education programs and, unless academic performance is adequate, other criteria do not matter.

Many educators have been troubled by the fact that the predictive limits of the admissions process are often faculty of education course grades and associate teachers' ratings. Queen's Faculty of Education has sought to extend this limit. This raises some problems. Whereas everyone will agree that a good student must at least have good grades, there is little agreement about the characteristics of a good teacher. Nonetheless, we offer a simplistic set of results directed at identifying success as a teacher.

The first criterion for success is "years of teaching since graduation". This is a baseline criterion widely accepted in industry and is usually called "seniority". This is a weak measure of "coolness", but one cannot evaluate other criteria unless the person is employed. The second criterion of post-graduate performance is the principal's rating of the teacher's performance. This is a simple year-by-year rating of teacher performance on a scale from poor to excellent. By examining the admissions and teaching characteristics of rated teachers, we may begin to infer what characteristics make a good teacher, at least from the principal's viewpoint.

Data collected in this study were 255 principals' ratings on teachers employed in their schools and employment data on 346 teachers. These success measures were related to performance in the admissions interview and in practice teaching. Many statistically significant correlations

were found, but none larger than 0.33. The majority of significant correlations, however, were about 0.15 ± 0.04 . Exact values for the correlations are presented in Tables 15 to 19. We first examine the admission interview ratings in relation to seniority.

The teachers with the most teaching experience now were those that faculty had perceived as cheerful and cooperative in nature. They appeared alert, self-reliant, and verbally fluent in their interviews. Teacher ratings of this group were different. It was those who were least self-possessed and least responsive to the perceptions of the teachers that have obtained the most teaching experience. In contrast, administrators saw the currently more senior teachers as the more authoritative, purposeful and self-possessed of the interviewed applicants. The student view focussed upon those who were least helpful, but most resourceful and adaptable. The significant characteristics which predicted seniority seem to vary with the position of the rater. Each rater group has different concerns. Quite possibly, the characteristics found significant for each rater are those of most concern or most characteristic of people in those positions.

In addition to the admissions interview, the sample teachers were rated during their associate teaching experience. Those who were deemed most responsible and who made the best overall impression in their practice teaching, are those who have obtained the most seniority since graduation.

The other success criterion we considered was a principal rating of the teachers' performance in the current and previous school years. Many more variables were significantly related to this more subjective criteria. In part, this is because ratings tend to correlate better with other ratings than with alternative types of success measures (e.g., rank orderings, percentiles, etc.) Nonetheless, there were a variety of admission ratings which weakly predicted the 5 year post-graduation performance ratings of teachers made by principals.

For the faculty admission interview ratings of applicants, four of the characteristics which predicted seniority also predicted performance ratings. Specifically, those rated as cooperative, self-reliant, alert and fluent, were most likely to obtain good performance ratings, as well

Table 15

A. RATINGS GIVEN BY FACULTY MEMBERS
AT THE ADMISSION INTERVIEW

<u>Characteristic</u>	<u>Seniority</u>	<u>Performance Rating</u>
Responsive	.03	.07
Stimulating	.05	.07
Systematic	.03	.07
Communicative	.05	.12 *
Authoritative	.09	.01
Adaptable	.01	.08
Democratic	.01	.06
Purposeful	.03	.07
Cheerful	.16	.06
Responsible	.0002	.06
Personal Magnetism	.09	.05
Trusting	.02	.01
Self-Possessed	.07	.15 *
Open-Minded	.06	.02
Challenging	.05	.09
Kindly	.08	.02
Sense of Humour	.08	.09
Pleasant Voice	.09	.12 *
Vigorous	.05	.08
Consistent	.10	.16
Helpful	.08	.10
Resourceful	.05	.15
Expressive	.06	.12
Cooperative	.13 *	.13 *
Self-Reliant	.12	.21
Original	.02	.21
Rational	.05	.14
Alert	.19	.15
Fluent	.19	.15
Forceful	.03	.15

Table 16

B. RATINGS GIVEN BY STUDENTS AT THE
ADMISSION INTERVIEW

<u>Characteristic</u>	<u>Seniority</u>	<u>Performance Rating</u>
Responsive	.10	.004
Stimulating	.04	.12
Systematic	.04	.06
Communicative	.05	.07
Authoritative	.05	.07
Adaptable	.16 *	.07
Democratic	.07	.16
Purposeful	.04	.10
Cheerful	.09	.25
Responsible	.03	.006
Personal Magnetism	.05	.10
Trusting	.07	.03
Self-Possessed	.12	.05
Open-Minded	.13	.12
Challenging	.01	.10
Kindly	.05	.18
Sense of Humour	.10	.14
Pleasant Voice	.07	.18
Vigorous	.03	.07
Consistent	.04	.08
Helpful	.18	.02
Resourceful	.15	.04
Expressive	.05	.13
Cooperative	.05	.02
Self-Reliant	.008	.06
Original	.06	.07
Rational	.08	.11
Alert	.03	.08
Fluent	.07	.05
Forceful	.04	.02

Table 17

C. RATINGS GIVEN BY TEACHERS AT THE
ADMISSION INTERVIEW

<u>Characteristic</u>	<u>Seniority</u>	<u>Performance Rating</u>
Responsive	.23	.09
Stimulating	.07	.19 *
Systematic	.05	.12
Communicative	.16	.25 *
Authoritative	.06	.15
Adaptable	.15	.12
Democratic	.08	.11
Purposeful	.005	.28 *
Cheerful	.12	.02
Responsible	.004	.25 *
Personal Magnetism	.03	.14
Trusting	.16	.17
Self-Possessed	.25	.04
Open-Minded	.07	.04
Challenging	.07	.15
Kindly	.11	.10
Sense of Humour	.12	.09
Pleasant Voice	.07	.15
Vigorous	.04	.002
Consistent	.11	.08
Helpful	.05	.08
Resourceful	.06	.16
Expressive	.02	.20
Cooperative	.01	.12
Self-Reliant	.18	.16
Original	.04	.10
Rational	.16	.02
Alert	.12	.05
Fluent	.08	.39
Forceful	.06	.07

Table 18

D. RATINGS GIVEN BY ADMINISTRATORS
AT THE ADMISSION INTERVIEW

<u>Characteristic</u>	<u>Seniority</u>	<u>Performance Rating</u>
Responsive	.03	.06
Stimulating	.01	.06
Systematic	.07	.07
Communicative	.06	.006
Authoritative	.20 *	.17
Adaptable	.17	.06
Democratic	.08	.05
Purposeful	.24 *	.18
Cheerful	.06	.05
Responsible	.18	.08
Personal Magnetism	.01	.09
Trusting	.06	.03
Self-Possessed	.20	.32
Open-Minded	.02	.07
Challenging	.02	.10
Kindly	.07	.07
Sense of Humour	.04	.02
Pleasant Voice	.07	.03
Vigorous	.005	.07
Consistent	.0008	.12
Helpful	.07	.02
Resourceful	.13	.08
Expressive	.09	.16
Cooperative	.14	.03
Self-Reliant	.17	.16
Original	.03	.03
Rational	.04	.04
Alert	.08	.07
Fluent	.03	.02
Forceful	.09	.10

Table 19

E. RATINGS GIVEN BY SUPERVISING TEACHER AFTER
FIRST PRACTICE TEACHING ASSIGNMENT

<u>Characteristic</u>	<u>Seniority</u>	<u>Performance Rating</u>
Responsive	.002	.13 *
Stimulating	.04	.14
Systematic	.007	.13 *
Communicative	.06	.04
Authoritative	.04	.005
Adaptable	.005	.08
Democratic	.06	.17 *
Purposeful	.04	.10
Cheerful	.07	.07
Responsible	.16 *	.07
Personal Magnetism	.009	.09
Trusting	.04	.01
Self-Possessed	.10	.003
Open-Minded	.03	.009
Challenging	.03	.10
Kindly	.008	.06
Sense of Humour	.01	.004
Pleasant Voice	.06	.01
Vigorous	.01	.16 *
Consistent	.07	.11 *
Helpful	.04	.12 *
Resourceful	.05	.10
Expressive	.04	.10
Cooperative	.01	.09
Self-Reliant	.02	.03
Original	.009	.10 *
Rational	.03	.04
Alert	.07	.15 *
Fluent	.03	.004
Forceful	.03	.09

as obtaining the greatest seniority. In addition, good performance ratings were earned by those who were rational, consistent and self-possessed. Being communicative, expressive and having a pleasant voice also characterized the successful applicants. In addition, these successful teachers were rated as resourceful, original and forceful in faculty ratings of their admission interview performance. More generally, those making the best impression upon their school principals were those that made the best overall impression on faculty members at their admission interview.

In addition to impressing faculty, these applicants also made the best overall impression upon students rating them at the admission interview. Students perceived the subsequently successful teachers as more kindly and having a more pleasant voice than teachers who achieved lower post-graduation performance ratings.

Administrator ratings at the admission interview yielded only one characteristic which related to performance ratings, that is, self-possession. This was the only characteristic rated by this group with any significant degree of predictive power.

The teachers who conducted admission interviews had different perceptions than the other three rater groups, that is, the characteristics that made for their subsequently successful teacher had a different orientation. For them, the successful applicant was a responsible person, one who was purposeful and stimulating, as well as being communicative.

Surprisingly though, none of these characteristics were found significant in the practice teacher ratings. The characteristics which predicted later performance ratings changed completely in the classroom setting. The successful teacher was the one rated most helpful and most responsive in practice teaching. They were alert, original, and vigorous in their work. They were systematic and consistent while being democratic. Basically, the teachers obtaining the best principal ratings were those who made the best overall impression in their practice teaching experience. Interestingly enough, they were also the students who achieved the highest grades in their other faculty of education courses. It would

seem that the characteristics which make for a good student and a good practice teacher, also help make for a good teacher in subsequent practice.

These serious weak correlations cannot be easily summarized. The positive correlations can be explained "after the fact" with little difficulty, but the three negative correlations are counter-intuitive and puzzling. There are several explanations for observing these negative values, as well as for the small correlations we found in general. The most plausible of these explanations results from the very selected nature of the group we are observing. This resulting restriction of range phenomenon will be discussed after presentation of the remaining results.

For the moment we change the focus of the analysis. In previous analyses, we were looking for predictors of seniority and performance ratings among practicing teachers. These are acceptable success criteria according to the principles of organizational psychology, but we have not considered a more basic success criterion, i.e., getting a job in teaching. This is certainly a distinction which divides the graduates in a meaningful way. It is recognized that there may be many reasons for succeeding or failing to be employed five years after graduation which our analyses will not reveal. For example, where one wants to live will have a significant impact on employability, yet this is not captured in our analyses. Additionally, subject specialty area can have a major impact. Despite our cognizance of additional major factors, we have limited ourselves to analyzing the variables in the Queen's interview schedule.

Table 20 gives the average admission ratings of the currently employed and unemployed teachers. The average admission interview ratings which faculty members gave the currently employed group were higher on 24 of the 30 rating scale dimensions, but they were significantly higher for only two dimensions: dependent versus self-reliant and unintelligible versus fluent. There was no theoretical basis for expecting these two particular dimensions to be the only ones which were significant. It is probable that these are significant by chance alone, but there was a tendency for currently employed teachers to have obtained better admissions interview ratings from faculty members than did those who are currently unemployed.

TABLE 20- ADMISSION INTERVIEW RATINGS

CHARACTERISTIC	RATED BY FACULTY		RATED BY STUDENT	
	Average Employed	Average Unemployed	Average Employed	Average Unemployed
				s.d.
Responsive	5.32	5.24	5.48	5.40
Stimulating	4.85	4.91	4.89	5.05
Systematic	5.18	5.06	5.21	4.90
Communicative	5.05	4.93	4.83	4.92
Authoritative	4.97	4.87	4.82	4.58
Adaptable	5.14	5.16	5.03	5.08
Democratic	5.11	5.16	5.08	5.14
Purposeful	5.29	5.14	5.37	5.11
Cheerful	5.28	5.12	5.20	5.34
Responsible	5.46	5.38	5.40	5.38
Personal Magnetism	4.81	4.81	4.74	4.78
Trusting	5.31	5.22	5.15	5.29
Self-possessed	5.27	5.16	5.09	4.99
Open-minded	5.25	5.21	5.14	5.20
Challenging	4.77	4.74	4.74	4.57
Kindly	5.23	5.22	5.36	5.47
Sense of humour	5.04	4.95	4.87	5.11
Pleasant voice	5.08	5.10	5.12	5.33
				s.d.
				0.73
				1.15
				1.00 *
				1.12
				1.10
				1.07
				0.93
				0.91 *
				0.95
				0.84
				1.26
				0.85
				1.06
				0.97
				1.17
				0.84
				1.05
				0.99

TABLE 20 CONTINUED

CHARACTERISTIC	RATED BY FACULTY			RATED BY STUDENT		
	<u>Average Employed</u>	<u>Average Unemployed</u>	<u>s.d.</u>	<u>Average Employed</u>	<u>Average Unemployed</u>	<u>s.d.</u>
Vigorous	4.99	4.90	1.07	4.83	4.80	1.12
Consistent	5.37	5.24	0.87	5.28	5.05	0.91
Helpful	5.44	5.34	0.76	5.35	5.30	0.93
Resourceful	4.99	4.90	1.06	4.89	5.00	1.14
Expressive	4.99	4.94	1.15	4.87	4.98	1.13
Cooperative	5.51	5.41	0.76	5.34	5.51	0.85
Self-reliant	5.29	5.07	0.97 *	5.19	5.02	1.06
Original	4.83	4.74	1.02	4.57	4.66	1.19
Rational	5.42	5.32	0.74	5.37	5.24	0.78
Alert	5.22	5.30	0.92	5.33	5.39	0.81
Fluent	5.20	4.93	1.02 *	5.12	5.23	0.94
Forceful	4.83	4.66	1.24	4.56	4.44	1.29
Overall impression	5.14	5.04	1.01	5.03	5.18	1.04
Maximum sample size	337	222	559	171	94	265

*t-test with pooled variance estimate is significant at $p \leq 0.05$

As one can see, however, this was not true for the interview ratings given by the student interviewers. Students rated the employed group higher on half the characteristics, and lower on the remaining half. Overall, though, they tended to perceive the currently employed group in a slightly more negative way. Despite this perception, the only two significant differences observed were both in favour of the employed group. The employed group was rated as significantly more systematic and purposeful in their interview performance. These skills could certainly yield a slight advantage in "job-hunting" after graduation. Perhaps it is the ability to find work rather than the ability to teach which these ratings reflect.

Table 21 provides a contrast to these results. The raters in both cases here are practicing teachers, but the first set of ratings was based upon an admission interview and the second set upon observing the student teacher in a classroom setting. The contrast between these settings was marked.

The teacher-based interview presented a pattern similar to the other interview ratings, however, there were significant differences between comparison groups in any context. Even the overall impression of the two groups was similar.

On the basis of practice teaching observations, however, the two groups were significantly different on a number of characteristics. For all 17 significant characteristics, the currently employed group was rated more positively than the unemployed group. In fact, for 29 out of 30 comparisons, the employed group was rated more positively. A comparison of the grades of currently employed (3,679) versus unemployed (3,647) graduates also revealed a significant difference in favour of the employed group ($t = 2.09$, $p .05$). In this case, and in the others, the statistically significant differences were not large enough to be truly diagnostic. Nonetheless, they suggested a number of possibilities and lead us to raise a number of substantive issues.

TABLE 21: PERSONALITY RATINGS BY TEACHERS

CHARACTERISTIC	RATED AT ADMISSION INTERVIEW			RATED IN ASSOCIATE TEACHING		
	Average Employed	Average Unemployed	s.d.	Average Employed	Average Unemployed	s.d.
Responsive	5.23	5.25	0.89	5.17	5.04	1.03 *
Stimulating	4.89	4.90	1.10	4.72	4.51	1.16 *
Systematic	5.05	4.95	1.04	4.95	4.84	1.12
Communicative	4.84	4.80	1.17	4.68	4.58	1.18
Authoritative	4.66	4.62	1.16	4.65	4.43	1.13 *
Adaptable	5.09	5.06	0.99	5.09	5.07	1.05
Democratic	5.12	5.04	0.99	5.09	5.06	0.99
Purposeful	5.17	5.14	1.01	5.10	4.98	1.05 *
Cheerful	5.13	5.25	0.95	5.20	5.04	1.10
Responsible	5.29	5.19	0.95	5.27	5.17	1.05
Personal magnetism	4.70	4.76	1.21	4.69	4.49	1.23 *
Trusting	5.25	5.09	0.94	5.24	5.19	0.89
Self-possessed	4.98	4.80	1.12	5.11	4.99	1.08
Open-minded	5.24	5.09	0.97	5.23	5.15	0.98
Challenging	4.72	4.59	1.16	4.61	4.48	1.22
Kindly	5.23	5.22	0.99	5.29	5.20	0.92
Sense of humour	4.87	4.93	1.04	4.98	5.00	1.07
Pleasant voice	5.10	5.13	1.01	5.02	4.98	1.07

CHARACTERISTIC	RATED AT ADMISSION INTERVIEW			RATED IN ASSOCIATE TEACHING		
	<u>Average Employed</u>	<u>Average Unemployed</u>	<u>s.d.</u>	<u>Average Employed</u>	<u>Average Unemployed</u>	<u>s.d.</u>
Vigorous	4.83	4.87	1.08	4.80	4.55	1.15 *
Consistent	5.17	5.09	0.96	5.25	5.13	0.92 *
Helpful	5.31	5.35	0.74	5.46	5.28	0.79 *
Resourceful	4.96	4.81	1.15	4.93	4.77	1.16 *
Expressive	4.85	5.03	1.11	4.80	4.60	1.14 *
Cooperative	5.37	5.41	0.79	5.49	5.35	0.83 *
Self-reliant	4.97	4.82	1.18	5.05	4.79	1.18 *
Original	4.75	4.71	1.03	4.74	4.53	1.13 *
Rational	5.30	5.22	0.79	5.40	5.22	0.83 *
Alert	5.26	5.30	0.95	5.24	5.03	1.01 *
Fluent	5.19	5.15	0.90	5.20	5.08	0.95 *
Forceful	4.68	4.63	1.30	4.59	4.28	1.21 *
Overall impression	5.00	5.01	1.03	5.04	4.86	1.00 *
Maximum sample size	196	132	328	765	509	1274

*t-test with pooled variance estimate is significant at $p < 0.05$

First, they suggest the obvious fact that an admission interview is vastly different from an associate teaching experience. In the interview, the raters must operate on superficial impressions of the student, and they must try to imagine how good a teacher this person would make. In practice teaching, the rater has a chance to actually observe the person's competence, not just to think of how good they might be.

It is axiomatic, then, that stronger opinions should be formed in the classroom setting. That these classroom ratings predict who will later be employed is not surprising for two reasons. First, a sample of two persons' work on day 10 is the best predictor of their work on day 100. Secondly, the classroom experience is closer in time than the admission interview to our study observations. As a result, fewer factors are likely to have intervened and distorted pre-existing relationships.

Two simple inferences can be made from observed employed-unemployed contrasts and from correlations between success criteria and earlier performances. First, those who were hired were the best of the students and, thus, only the best members of the class are presently employed. Secondly, of those who are employed, the most successful are those who were the most successful students. That is, those who received the best grades and the best associate teacher ratings are most successful at getting jobs, retaining jobs, and being considered good at those jobs.

This is a very appealing conclusion but is probably too simplistic. It seems equally likely that the people who obtained jobs were those who performed better in the Queen's program. This is to say that when school boards hired teachers, they tended to hire those with the best academic grades and best practice teacher ratings. Certainly this is to be expected, and explains the observed differences somewhat more satisfactorily. In short, admission interviews do not predict later employment because they are not a basis for hiring. Academic grades and associate teacher ratings predict later employment and it is here assumed that these factors are the basis for employment decisions.

In the final analysis we do not know definitively whether the employed teachers are any better than the unemployed group. The issue hinges entirely upon the validity of the rating scale and particularly upon what rating differences are meaningful. In order to validate the instrument we must have some criteria for what makes a good teacher.

Our first assumption was that principals are in a position to globally assess teacher effectiveness and that their assessment of teachers' effectiveness is indeed valid. Without explicit criteria to define attributes of good teachers, one is left with using data assumed to relate positively to "good teaching". An argument could be advanced that such evidence is circular in nature and thus misses the essential focus of what such a study should focus upon. Hence, without an explicit construct defined as to what characterizes a "good teacher" and an identification of variables that lend themselves to reliable measurement, one may overlook relevant data.

As an addendum, we would like to raise yet another possibility. It is possible that real differences do exist between the employed and unemployed group. The reason that these may go undetected is "rating-inflation". Raters have a perverse tendency these days to want to rate everyone positively. Thus an average, a good, or an excellent performer is rated a six or a five. The rest of the scale goes unused and our ability to distinguish between qualities of performance is lost. We offer no practical solution to this problem other than a request for more meaningful ratings. Whether this will help us to distinguish good from bad in teaching, however, is doubtful. It is only by establishing standards of practice that we can really begin to tackle this issue. As it stands now, our only standard is seniority.

These comments have by-passed consideration of the correlational results. These cannot be explained in terms of hiring criteria, but the weakness of these relationships suggest two alternative statistical explanations.

Using seniority as a success criterion, 15 correlations were found significant. By chance alone, statistical theory predicts that eight

correlations will be significant, even though there are no true differences in groups underlying this observed significance. Considering that this data consists of five independent comparisons of the same 31 variables, and that only one variable (self-possessed) was significant in more than one comparison, it seems likely that chance has played a major role in the results obtained in these analyses.

The other phenomenon of relevance here is restriction of range or pre-selection. The use of correlational or variance-based analysis methods is predicated upon the assumption that all variables are normally distributed in the population of interest. Each time we eliminate some people from further consideration, this assumption of normality becomes less tenable. Consider the selection process for this group: A% of high school students apply and are selected for admission to university; B% of this A% obtain degrees; C% of the degree group apply and are accepted to the faculty of education; D% of this group graduate; E% obtain jobs. In the original case (high school students), the A% had a relatively normal distribution of abilities and there were a lot of differences between applicants. The A% had different career interests and goals and attended a wide variety of schools. By the time we reach our final stage, however, the group is much narrower. Now we are referring to people with enough ability and interest to get a Bachelor's degree. This is not a very large proportion of the population at large. In addition, we are only referring to degree holders interested in attending a single B.Ed. program in Kingston, Ontario. Thus the group is further homogenized by their common interest in teaching and their willingness to attend a program in Kingston. This group is unlikely to vary widely in ability or interest compared to the group we started with. As a result of this lack of variation, we observe very weak correlations.

Given that these people are fairly homogenous, it may well be that one graduate is as intellectually capable of teaching as another graduate. If this is the case, then we could select teachers by lottery and do as well as we are doing now.

Before anyone gets too angry about this suggestion, let us clarify that we are not posing it as a real alternative. What we are doing instead

is suggesting the bankruptcy of the classical norm-referenced predictive approach for purposes of teacher selection and for purposes of identifying the successful teacher. What is required now are value oriented studies; studies where a group of educators begin to decide what constitutes effective teaching. The empirical approach cannot really establish what constitutes effective performance, unless some common definition of "effective" is operative.

Even this approach is too simple. The complete issue is what constitutes effective teaching under what circumstances. An effective science teacher may differ markedly from an effective English teacher. An effective English teacher in one neighbourhood may be ineffective in a neighbourhood of different social class or ethnic background. Alternatively, it may be that most teachers are effective. The task then is to identify those that are ineffective (under what circumstances) and to establish minimum competency standards.

The task then is not an easy one. Rather than predict, we must systematically provide the conceptualization to formulate the characteristics of a "good" teacher and define our educational needs. What should a teacher be able to do and how can we know if they are able to do it? We offer no easy answers, only a help in focus. An optimal situation would be to determine what a starting teacher must be able to do at time of hiring. Future needs cannot be easily anticipated without gross errors. Also, if current needs can be satisfied by existent skills, then other skills can be developed on a "need-to-know" basis.

Without solutions to current educational staffing problems, we may expect things to continue as they are. Jobs will be retained on the basis of seniority, and little new blood will enter the system at all. We will continue avoiding the issue of effectiveness, unless it implies that everyone in the system can retain their jobs. The consequences of this policy are unknown. The average age of teachers will continue to rise and it will increasingly differ from that of students. If identification with the teacher or modelling of teacher behaviour is an important force in education, then the quality of education will

gradually decline. If this is not the case, what is? How will new ideas work their way into the school system? How will a group accustomed to teaching by 1978 methods respond to the educational challenges of the 80's and 90's? How will the scarcity of educational jobs affect the quality of new teachers?

We raise this host of questions after considering the findings of the Queen's study. Only one thing is apparent. The answer to these questions will have a major impact on our society. It is up to us either to create our future through affirmative action or to find ourselves the victims of forces we don't yet understand. The decision depends upon the educators of this province.

REFERENCES

- Bolton, D. L. *Instructor's Guide for Use of Simulation Materials for Teacher Selection*. Columbus: University Council for Educational Administration, 1970.
- Bolton, D. L. *Selection and Evaluation of Teachers*. Berkeley: McCutchan, 1973.
- Brown, I., Weinstein, E. L., & Wahlstrom, M. W. *Admission and Selection Procedures for Nursing: Literature Review and Annotated Bibliography*. Toronto: The Ontario Institute for Studies in Education, 1978, 118 pp.
- Campbell, D. T., & Fiske, D. W. Convergent and divergent validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- Chauncey, H. & Frederiksen, N. The functions of measurement in educational placement. In E. F. Lindquist (Ed.) *Educational Measurement*. Washington, D. C.: American Council on Education, 1951, pp. 85-116.
- Cronbach, L. J. Test Validation. In R. L. Thorndike (Ed.) *Educational Measurement* (2nd ed.). Washington, D. C.: American Council on Education, 1971, pp. 443-507.
- Cronbach, L. J., & Gleser, G. *Psychological tests and personal decisions* (2nd ed.). Urbana: University of Illinois Press, 1965.
- Davis, J. A. Faculty perceptions of students:I. The development of the Student Rating Form. *Educational Testing Service Research Bulletin*, 1964, No.10.
- Davis, J. A. Faculty perceptions of students:VI. Characteristics of students for whom there is faculty agreement on desirability. *Educational Testing Service Research Bulletin*, 1966, No.28.
- Fishman, J. A. Unsolved Criterion Problems in the Selection of College Students. *Harvard Educational Review*, 1958, 28, 340-349.
- Gough, H. C., Hall, W. B., & Harris, R. E. Admissions procedures as forecasters of performance in medical training. *Journal of Medical Education*, 1963, 38, 983-998.
- Guilford, J. P. *Psychometric methods* (2nd ed.), New York: McGraw-Hill, 1954.

- Hills, J. R. Use of measurement in selection and placement. In R. L. Thorndike (Ed.) *Educational Measurement* (2nd ed). Washington, D. C.: American Council on Education, 1971, pp. 680-732.
- Hills, J. R., Gladney, & Klock, J. A. Setting cutoff scores in selective admissions. *Journal of Educational Measurement*, 1967, 4, 263.
- Hilton, T. L., & Myers, A. E. Growth study II. Personal background, experience, and school achievement: An investigation of the contribution of questionnaire data to academic prediction. *Journal of Educational Measurement*, 1967, 4, 69-80.
- Horst, P. A technique for the development of a differential prediction battery. *Psychological Monographs*, 1954, 68 (Whole No. 380).
- Horst, P. A technique for the development of a multiple absolute prediction battery. *Psychological Monographs*, 1955, 69 (Whole No. 390).
- Horst, P. *Psychological measurement and prediction*. Belmont, Calif.: Wadsworth, 1966.
- Kelly, E. L. Theory and techniques of assessment. *Annual Review of Psychology*, 1954, 5, 286-290.
- Klein, S. P., & Hart, F. M. The nature of essay grades in law school. *Educational Testing Service Research Bulletin*, 1968, No. 6.
- Lavin, D. E. *The prediction of academic performance*. New York: Russell Sage Foundation, 1965.
- Lindquist, E. F. An evaluation of a technique for scaling high school grades to improve prediction of college success. *Educational and Psychological Measurement*, 1963, 23, 623-646.
- Linn, R. L. Grade adjustments for prediction of academic performance: A review. *Journal of Educational Measurement*, 1966, 3, 313-329.
- Lunneborg, P. W., & Lunneborg, C. E. The differential prediction of college grades from biographical information. *Educational and Psychological Measurement*, 1966, 26, 917-925.
- Ryans, D. G. *Characteristics of Teachers: Their Description, Comparison, and Appraisal*. Washington, D. C.: American Council on Education, 1960.

- Sawyer, J. Measurement and prediction :clinical and statistical.
Psychological Review, 1966, 66, 178-200.
- Sechrest, L. Incremental validity: A recommendation. *Educational and Psychological Measurement*, 1963, 23, 153-158.
- Smith, M. B. Explorations in competence: A study of Peace Corps teachers in Ghana. *American Psychologist*, 1966, 21, 555-566.
- Weinstein, E. L., Brown, I., & Wahlstrom, M. W. *A Review of Admission and Selection Procedures for Diploma Nursing Programs in Colleges of Applied Arts and Technology: Program Survey and Recommendations*. Contract Research report submitted to the Ontario Ministry of Colleges and Universities, 1976, 125 pp.

APPENDICES

APPENDIX I

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

APPLICANT (CHECK ONE)

☐ TAMI BECK
☐ CAROL BROWN
☐ PATRICIA HANSEN
☐ JUDY OLSON
☐ CONNIE ANDREWS
☒ TERRY COLLINS
☐ MELODY COCHRAN
☐ NINA TAYLOR
☐ JEAN MILLER
☐ JANET LEE
☐ MARLENE WEST
☐ SALLY MICHELSON
☐ MARY ABRAHAMS
☐ LOIS WILLIAMS
☐ SHARON BARNES
☐ PAULA LEWIS

IMPORTANT ATTRIBUTES

RANK

ORGANIZATION 1 2 3 4 5
 SOCIABILITY 1 2 3 4 5
 ORIGINALITY 1 2 3 4 5
 EMPATHY 1 2 3 4 5
 BUOYANCY 1 2 3 4 5

INTERVIEWER'S NAME

DATE

A. PLANNING AND ORGANIZING CLASSWORK

purposeful : aimless
 systematic : disorganized
 cooperative : antagonistic
 original : unimaginative

B. CLASSROOM MANAGEMENT

punctual : tardy
 controlled : disorderly
 consistent : inconsistent
 flexible : fixed
 fair : partial
 responsive : indifferent

C. CREATING A MOTIVATIONAL ENVIRONMENT

steady : spasmodic
 flexible : rigid
 broad-minded : narrow-minded
 sensitive : unfeeling
 kindly : critical
 sense of humor : humorless
 stimulating : dull

D. INSTRUCTION

alert : apathetic
 resourceful : inflexible
 poised : agitable
 helpful : hindering
 inspirational : uninspiring
 precise comm. : fuzzy
 pleasant : harsh

E. EVALUATION

continuous : erratic
 rational : irrational
 systematic : disorganized
 just : inequitable

F. GUIDANCE AND COUNSELING

resourceful : trite
 patient : impatient
 approachable : formal
 inspiring : uninspiring
 communicative : uncommunicative

G. OUT-OF-CLASSROOM PROFESSIONAL ACTIVITIES

active : evading
 responsible : irresponsible
 skillful : unskillful
 accurate : inexact
 punctual : tardy
 constructive : antagonistic
 progressive : stagnant

H. RELATIONS WITH STAFF AND PARENTS

approachable : aloof
 cooperative : uncooperative
 discreet : imprudent
 responsible : irresponsible
 open-minded : narrow minded
 effective : ineffective

J. SCHOOL-COMMUNITY RELATIONS

active : inactive
 initiates : follows
 effective : ineffective

SUPPLEMENTARY INFORMATION

1. ORAL ENGLISH

Please rate the candidate's proficiency in speaking the English language:

	1	2	3	4	5	6	7	
Low proficiency	64	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	High proficiency

If the candidate scored less than 4, describe in a few words the nature of the shortcomings.

2. COMPETENCE IN OTHER LANGUAGES:

Please ascertain the candidate's competence in other languages, indicating whether it is in the spoken or written form (or both) and estimating the level of proficiency.

3. OTHER SUPPORTING TALENTS.

Indicate whether the candidate has skills or talents that might be useful for a teacher, e.g. music, art , sports, etc.

4. OTHER COMMENTS

PROCEDURES FOR INTERVIEWINGA. GENERAL GUIDE-LINES

1. Each interviewing panel will be headed up by a McArthur College faculty member who will act as chairman, and in that capacity serve as the university's official representative in all matters pertaining to the work of the interviewing team.
2. In keeping with the College's commitment to humane processes, interviewers will seek to make the applicant as comfortable as circumstances allow and endeavour to minimize the level of anxiety and stress during the interview.
3. Since a number of applicants may fail to earn admission to McArthur as a result of their interview ratings, it is vital that all interviewees be accorded every courtesy and consideration in order for them to make as favourable an impression as possible and to leave with the feeling that they were given a sympathetic hearing.

One specific illustration of the foregoing principle is to provide the applicant with an opportunity, toward the conclusion of the interview, to relate anything significant about himself that may not have been touched on during the discussion up to that point.

4. Although McArthur representatives will attempt to respond to any questions asked by the applicant, they will bear in mind that "the individual who does the talking is actually the person being interviewed". Therefore, they will try to structure the discussion so that the applicant, in fact, does most of the speaking during the interview.

5. Interviews will normally last 20 minutes, allowing another 10 minutes for the necessary evaluative and administrative activities.
6. Interviewing teams are to proceed on the assumption that all interviewees have the necessary academic qualifications for admission. The interview need not, indeed should not, be focussed on evaluating the candidate's subject area competence.
7. Only the chairman of the interviewing panel will have access to candidates' application forms, but he is encouraged to share with his fellow-interviewers relevant information about the candidates that may be helpful in the conduct of the interviews. For instance, a reference on the Personal Data Sheet to summer work with youngsters may serve as a point of entry into a discussion about the candidate's leadership experiences.
8. In the case of some applicants, it will assist decision-making by the Admissions Committee if there were an indication of a third Curriculum choice in the event that their second option choices are closed. Chairmen are asked to offer an opportunity to every candidate to identify such a third option: this choice ought then to be recorded on the application form.
9. Certain precautions are to be observed in the use of the Interview Assessment sheets:

- (1) Candidates are not to be shown the Interview Assessment sheets;

- (2) Although interviewers should feel free to jot down notes during the interview, they will complete the summary form itself only after the candidate's departure.
- (3) It is critically important that interviewers arrive at their own judgments independently of each other; therefore, interview summaries ought to be completed prior to any comparison of impressions.

B. INSTRUCTIONS Re INTERVIEW ASSESSMENT FORM

1. Interviewers are asked to make certain that they provide all the data requested at the top of the Assessment form, regarding the applicant and the interviewer. The key-punch operators will require this information to do a satisfactory job in the preparation of computer cards.
2. The thirty bipolar items under Personal Dimensions relate to the six important personal/professional attributes (listed in the following section on the form). The five blocks of six items each were set up for purposes of readability alone and are not organized on any topical basis.

Since individuals' future careers may depend on the interviewers' objectivity and thoroughness, the following guide-lines ought to be observed:

- (1) Rate each dimension independently. Avoid a syndrome approach which results in pattern responses. (Particularly when fatigued, resist the tendency toward uniform rating on all dimensions.)
- (2) Use the full scale. All seven intervals in the continuum are to be considered. The end-points are to be utilized, and the mid-point is not to be over-utilized.
- (3) Our advice from measurement experts is that interviewers should be urged to attempt to complete every item if at all possible. It is crucial to have the data quite complete.

The foregoing instructions have been intentionally expressed in a very positive, pointed way so as to emphasize the need for interviewers to be discriminating and decisive.

But a word of reassurance may now be in order. The interview assessment instrument deals with approximations and does not call for agonizing appraisal by the rater. Remember that your individual rating of any particular dimension constitutes a relatively small element in the total admissions procedure. So do your careful best, and then relax.

3. The section titled Order of Important Attributes asks interviewers to rank-order the six major factors, from 1 to 6 with no ties. Since this section was designed to provide internal validation in subsequent research, interviewers can assume that their responses here will not affect decision-making about admission.
4. The Comments portion is optional and provides an opportunity for the inclusion of important information not covered in previous sections or in the application form.

C. SUGGESTIONS Re INTERVIEWING STRATEGY

Many of the specific qualities included in the interview summary form will be manifested rather naturally during the conversation and will require no special strategy. Qualities of that sort include expressiveness, sense of humour, and so forth.

But other characteristics may have to be ascertained by inference as the result of rather careful questioning. Attributes such as empathy, leadership, originality and organizational ability will require that kind of skilful, indirect approach.

To assist interviewing panels in the conduct of their discussions it is suggested that most interviews will need to honour certain natural stages. Although specific questions and topics of discussion will vary from case to case, the general design of most interviews will incorporate these features.

STAGE ONE - REDUCTION OF TENSION

- (1) Introductions and explanations
- (2) Casual conversation
- (3) Identification of a third Curriculum option

STAGE TWO - LEADERSHIP QUALITIES

- (1) Invite the candidate to describe some leadership experiences that he has had.
- (2) Then ask the candidate to recall some problems or difficulties he encountered. e.g. What was the hardest thing you had to handle as a member of the Students' Council?

- (3) Finally, to assess qualities such as decisiveness and assumption of responsibility, request the candidate to tell how he coped with the situation.

e.g. Well, as camp counsellor what did you do in that emergency?

STAGE THREE - ORIGINALITY

In order to gauge the candidate's capacity for flexible, imaginative thinking, it is important to avoid topics where simple recall or resort to conventional wisdom will suffice. There emerges a need for the novel and the unanticipated, such as hypothetical situations.

e.g. How would you respond if your students ask to participate in the planning of their courses?

e.g. Suppose you're teaching Grades 7 and 8, and the kids don't like that particular unit.

e.g. (For a non-Math candidate) How do you think Mathematics ought to be taught in High School?

e.g. Describe how you would like to organize your own classroom's physical arrangements.

e.g. What alternatives are there to traditional examinations and grading?

e.g. Suppose the Department Head has imposed a course outline, and it won't go with the kids in your class.

e.g. What steps would you take if you discover that you have an inadequate background in one of your teaching fields?

e.g. There are two students talking at the back of the class. What alternative courses of action are open to you as their teacher?

e.g. Suppose the students in your school don't seem to care about passing....

STAGE FOUR - EMPATHY

Here again the interviewers may aim at illustrations from previous experience or responses to hypothetical problems.

Illustration of the former:

e.g. How did you get along with your room-mates at the co-op?

e.g. How did you resolve the interpersonal difficulty with your supervisor at work?

Examples of the latter:

e.g. Are students more difficult to discipline to-day than they used to be?

e.g. Suppose you have a "multi-problem" kid in your class (inattentive, sleepy, failing, needs glasses, works nights in the bowling-alley, father is crippled, a large family, on welfare), how would you as the teacher respond?

e.g. Suppose a kid in your class gets beaten up on the way home from school by some of his class-mates.

e.g. You suspect some students in your class are on drugs....

STAGE FIVE - ORGANIZATIONAL ABILITY

e.g. How did you make your career-choice?

e.g. How did you happen to choose your university?

e.g. What are your reasons for attending this
McArthur interview?

e.g. What are some of the necessary stages in the
organizing of a field trip?

e.g. What are the principal steps in applying for a
teaching position?

STAGE SIX - CONCLUSION

- (1) Is there anything else that you would like
to say about yourself that we haven't touched on?
- (2) Clarification of next steps in admissions procedures
- (3) Good-byes

It will be understood that the stages identified
above will not necessarily follow each other in the same
consecutive order as shown here.

A final word of advice: it would make good sense
for the three members of the panel to spend about an hour
together going over these guide-lines and agreeing on a
mode of operation.

APPENDIX 4

INFORMATION Re TEACHING ASSESSMENT FORM

BACKGROUND

This is a word of explanation about the long yellow sheets. Briefly, we are conducting a follow-up study on this year's candidates to see how well our selection procedures predicted their success as teachers. All of our associate teachers and faculty supervisors are being asked to complete one form for each candidate whose teaching performance they observe in the schools.

Candidates are not to be shown your assessment of their performance on the Teaching Assessment form. Your ratings on this instrument will have no bearing whatever on candidates' academic grades.

GUIDE-LINES

1. Simply ignore Student Number and Supervisor No.
2. The thirty bipolar items under Personal Dimensions relate to six important attributes: Empathy, Organization, Leadership, Buoyancy, Originality, and Professional Impression. The five blocks of six items each were set up for purposes of readability alone and are not organized on any topical basis.
3. Rate each dimension independently. Avoid a syndrome approach which results in pattern responses.
4. Use the full scale. All seven intervals in the continuum are to be considered. The end-points are to be utilized, and the mid-point (the average) is not to be over-utilized.
5. Our advice from measurement experts is that interviewers should be urged to attempt to complete every item if at all possible. It is crucial to have the data quite complete.
6. The section titled Order of Important Attributes asks interviewers to rank-order the six major factors, from 1 to 6 with no ties. Frequently, this is a difficult section to complete. (It was designed to provide internal validation in subsequent research.)

7. The Comments portion will ordinarily be left blank.
8. By way of reassurance, the assessment instrument deals with approximations and does not call for agonizing appraisal by the rater. So do your careful best, and then relax.

Associates may enclose completed Teaching Assessment forms along with the regular Associate Teacher's Report in the self-addressed envelopes from Professor Hennessy's office.

Please be assured that we do appreciate your assistance in this matter.

W. S. Peruniak
Assistant Dean

